Non-Parametric Volatility Measurements and Volatility Forecasting Models

by

CORNEL DU TOIT

Assignment presented in partial fulfilment of the requirements for the degree of
Master of Commerce
at the University of Stellenbosch

SUPERVISOR: Prof. W.J. Conradie

March 2005

Declaration

I, the undersigned, hereby declare that the work contained in this assignment is my own original work and that I have not previously in its entirely or in part submitted it at any university for a degree.

Abstract

Volatilty was originally seen to be constant and deterministic, but it was later realised that return series are non-stationary. Owing to this non-stationarity nature of returns, there were no reliable ex-post volatility measurements. Subsequently, researchers focussed on ex-ante volatility models. It was only then realised that before good volatility models can be created, reliable ex-post volatility measuremetns need to be defined.

In this study we examine non-parametric ex-post volatility measurements in order to obtain approximations of the variances of non-stationary return series. A detailed mathematical derivation and discussion of the already developed volatility measurements, in particular the *realised volatility*- and *DST* measurements, are given In theory, the higher the sample frequency of returns is, the more accurate the measurements are. These volatility measurements referred to above, however, all have short-comings in that the *realised volatility* fails if the sample frequency becomes to high owing to microstructure effects. On the other hand, the *DST measurement* cannot handle changing instantaneous volatility. In this study we introduce a new volatility measurement, termed *microstructure realised volatility*, that overcomes these short-comings. This measurement, as with realised volatility, is based on quadratic variation theory, but the underlying return model is more realistic.

Oorsig

Volatiliteit is oorspronklik as konstant en deterministies beskou, dit was eers later dat besef is dat opbrengste nie-stasionêr is. Betroubare volatiliteits metings was nie beskikbaar nie weens die nie-stasionêre aard van opbrengste. Daarom het navorsers gefokus op vooruitskatting-volatiliteits modelle. Dit was eers op hierdie stadium dat navorsers besef het dat die definieering van betroubare volatiliteit metings 'n voorvereiste is vir die skepping van goeie vooruitskattings modelle.

Nie-parametriese volatiliteit metings word in hierdie studie ondersoek om sodoende benaderings van die variansies van die nie-stasionêre opbrengste reeks te beraam. 'n Gedetaileerde wiskundige afleiding en bespreking van bestaande volatiliteits metings, spesifiek *gerealiseerde volatiliteit* en *DST*- metings, word gegee. In teorie sal opbrengste wat meer dikwels waargeneem word tot beter akkuraatheid lei. Bogenoemde volatilitieits metings het egter tekortkominge aangesien *gerealiseerde volatiliteit* faal wanneer dit te hoog raak, weens mikro-struktuur effekte. Aan die ander kant kan die *DST* meting nie veranderlike oombliklike volatilitiet hanteer nie. Ons stel in hierdie studie 'n nuwe volatilitieits meting bekend, naamlik *mikro-struktuur gerealiseerde volatiliteit*, wat nie hierdie tekortkominge het nie. Net soos met *gerealiseerde volatiliteit* sal hierdie meting gebaseer wees op kwadratiese variasie teorie, maar die onderliggende opbrengste model is meer realisties.

Acknowledgements

I would like to thank the following people for their help during the course of this study:

- My supervisor, Prof Willie Conradie, for his aid, enthusiasm and support.
- Prof Tertius de Wet for valuable comments that led to a substantial improvement of my thesis.
- My mother, Cynthia du Toit, for the time spent to edit my thesis.
- Dr. Alex Mendecki for allowing me the time off to work on my thesis.

Contents

1

Introduction

1.1 Background and motivation for study

In layman's terms volatility can be defined as the fluctuation over time of a financial random variable, but statistically it is **defined as the second moment of a financial random variable**. The volatility of asset returns is one of the key inputs in financial markets. It is used to determine the probability that specific returns are being achieved, which is crucial in the financial decision-making process. Asset allocation, the construction of the efficient frontier and the evaluation of derivatives depend heavily on volatility estimation. It is a well-known fact that asset returns are fairly fickle, whereas volatility is relatively predictable. The problem with volatility estimation however, is that volatility is unobservable. To date, there have been three stages in the evolution of volatility estimation.

During the first stage volatility was believed to be constant and deterministic. This assumption implies that stock returns are independently and identically distributed (i.i.d). Evidence against the independent assumption of stock returns was found in the pioneering investigation of the 1960's, where it became evident that large price movements tended to be followed by large price movements and conversely, small price movements tended to be followed by small price movements. It was also found that when calculating variances of returns on overlapping samples and non-equal length time-periods, the variances differed significantly. To measure and forecast volatility, it was clear that sample variances of returns should not be used. [See Black, 1960, Fama, 1965 and Joyce and Vogel, 1970 on this topic.]

The second stage was introduced in 1982 with the realisation that volatility itself is time-dependent and that the return series is non-stationary. **If a return series is non-stationary, then historical variances of returns cannot be used to measure volatility as they are not well-defined**. Volatility was considered as an unobservable variable. Historically, implied volatility that uses return models such as Black-Scholes, and indicators of volatility such as rolling daily squared returns were used as a volatility measurement to learn about the characteristics of volatility. This was done because volatility was seen as a latent (unobservable) process. The risk of using models to measure ex-post volatility is that there is the possibility of model misspecification. On the other hand, the variance of daily ex-post returns contains noise and the variance of the noise is too large relative to that of the signal. [See Anderson and Bollershev 1998.] If two years of daily data were to be used for calculating rolling daily squared returns, conditions may change within that time-period, and the older information may become irrelevant. However, should one use only a month of daily data, the estimator will have a high varaince due to few data points.

The main purpose during this stage was to predict ex-ante expected volatility using parametric models (such as the well-known ARCH-GARCH models) that incorporate the various characteristics learnt from the ex-post volatility measurements. In general these kinds of models suffer from two problems: firstly, they are not able to explain some empirical characteristics of financial data, and secondly, the estimation procedures are often rather complex. **The main problem during this stage was that a good volatility measement had not been developed, and this made the construction of good models difficult**. Even if the type of model had been correct, it could not be fitted or tested properly till a sensible volatility measurement were found.

During the 1990's researchers shifted their attention to ex-post volatility by using non-parametric approaches. The third stage was entered in 1998. In Anderson, Bollershev, Diebold and Labys (2001a,b), Barndoff-Nielsen and Shepard (2001), (2002a,b and c) and Comte and Renault (1998), **a model-free (non-parametric) volatility measurement was specified. It was termed realised volatility, and it satisfies the criteria of a good measurement**. Volatility could hence be seen as "observable" given this measurement. The fact that no model is specified and that the new measurement of volatility is relatively error-free (in the sense that it approximates the second moment of returns very well) are huge advantages that this measurement has over other methods. Once these issues regarding volatility were resolved, the characteristics of volatility (known as *stylised facts* see Section 2.4) could be more clearly identified and studied. Only then could an appropriate model be constructed and tested for forecasting purposes.

In this thesis three volatility measurements are to be examined. In addition to realised volatility, a measurement formulated by Corsi and Cursi (2003), is also to be examined. A model will then be suggested that can be used to forecast volatility. **Finally, a new volatility measurement, termed microsructure realised volatility, will be explained. This volatility measurement attempts to overcome some shortcomings of the other two measurements.**

1.2 Overview

Chapter 2 deals with ex-ante expected volatility using parametric models. The well-known ARCH-GARCH models are discussed. We also give some general problems that parametric models have as a volatility measurement.

Chapter 3 gives a parametric technique for estimating the drift- and diffusion component for a given stochastic differential equation.

Chapter 4 investigates a non-parametric volatility measurement, termed *realised voltility*, that is based on quadratic variation theory. As the result of this measurement, volatility can be seen as "observable" and further all the *stylised facts* of volatility could be observed for the first time.

Chapter 5 shows another volatility measurement, termed the *DST-measurement* that overcomes some of the sort-comings that *realised volatility* has. We also briefly explains *microstructure effects.*

Chapter 6 shows return- and volatility models for forecasting purposes. The volatility model caters for all the *stylised facts of volatility.*

Chapter 7 explains in detail the sort-comings that *realised volatility* and the *DST-measurement* have and gives a new volatility measurement, which we termed *microstructure realised volatility*, that overcomes these problems. This measurement is also based on quadratic variation theory but the return model is more applicable.

Chapter 8 shows the practical simulations. The superiority of the *microstructure realised volatility measurement* is shown.

A conclusion of this study is given in **Chapter 9**.

2

The Second Stage

2.1 Introduction

During the first stage it was assumed that return series are stationary. This implies that the return process has the same mean and variance at all times and therefore, the sample variance (called historical volatility) may be used for forecasting purposes. This measurement does not make use of current information to update estimates. The stationary assumption forms the basis of the random walk model. After studies had been made during 1960-1980, [See Black, 1960, Fama, 1965 and Joyce and Vogel, 1970] evidence was found against the random walk model assumption of stock returns. These studies revealed that volatility does change over time, that volatility clustering is present (large changes in returns tend to be followed by large changes and vice versa) and that volatility tends to rise when prices are falling, and to fall when prices are rising (known as the *leverage effect*). These characteristics (called *stylised facts*) became apparent in the studies by using implied volatility and volatility indicators. This led to the investigation into the non-stationary nature of return series. It was found that the time-dependency of volatility causes returns to be fat-tailed. The reasons for this are given in Chapter 6.

The second stage was introduced when the time-dependecy of volatility was modelled using parametric models. The main focus of these models was to explain ex-ante expected volatility. These models captured the above-mentioned stylised facts. The question was whether these volatility measurements could capture all the stylised facts, which only became apparent after the realised volatility measurement was specified. As it turned out, not all the the stylised facts were captured by the implied volatility and volatility indicators studies during the second stage.

Listed below are some of the models used during the second stage which are still in current use. These models imply that returns have fat-tails and hence can explain the empirical distribution of returns better than historical volatility does. The models below apply only to daily, or larger than daily, volatility and not to instantaneous volatility. In Chapter 3 instananeous volatility is discussed.

2.2 ARCH-type models

Let S_t be the market value of the share or index at time t. Let $p_t = \log(S_t)$ be the logarithmic price process. The continuous return process is therefore defined as $r_t = p_t - p_0$, with $t \in [0, T]$, i.e. $S_t = S_0 e^{r_t}$. Further, suppose that the return process can be modelled as:

$$r_t = \mu_t + \varepsilon_t,$$

where μ_t is the mean of r_t, and ε_t is the residual return, i.e. $r_t - \mu_t$.

It follows forthright that $E_{t-1}(\varepsilon_t) = 0$, where the subscript $t-1$ indicates that the expectation is taken at time $t-1$. The time index (days,weeks etc.) is for the individual to decide. Autoregressive Condional Heteroskedastic (ARCH) models are designed to eliminate the systematically changing variance of returns, resulting in the leptokurtosis in the distribution of returns. Consequently $E(\varepsilon_t^2) \equiv \sigma_t^2$ can be modelled.

2.2.1 The ARCH(q)- model

This model was adopted in a paper by Engle (1982) and incorporates volatility clustering. The ARCH(q)-model assumes that

$$\sigma_t^2 = \upsilon + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2, \tag{2.2.1.1}$$

where $\upsilon > 0$, and α_i, $i = 1, 2, ..., q$ with $\sum_{i=1}^{q} \alpha_i = 1$, are the constant linear declining weights.

The model in (2.2.1.1) gives larger weights to recent returns than it gives to older returns. ARCH-type models imply that large changes in returns tend to be followed by large changes, and vice versa. This, as mentioned earlier, is called volatility clustering and is one of the explanations why the i.i.d. assumption for returns does not hold. The model in (2.2.1.1) can cope with the time-varying nature of stock returns and models volatility as being conditional on past information.

2.2.2 The GARCH(p,q)-model

The Generalised ARCH-model, known as the GARCH-model, that was proposed in Bollershev(1986) is:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \, , \tag{2.2.2.1}$$

with $\omega > 0$.

The conditional (conditioned on past information) variance in (2.2.2.1) is a function of both past squared residual returns and past variances. The following condition must be satisfied for the model to make sense:

$$0 < \sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{q} \beta_j \leq 1 \, . \tag{2.2.2.2}$$

If the condition in (2.2.2.2) is satisfied, then the GARCH-model corresponds to an infinite ARCH-model, with the advantage that much fewer parameters need to be estimated. Also, if the condition in (2.2.2.2) holds, then the lowest possible variance assumed by this model is ω. A large number of empirical studies has found that a GARCH(1,1)-model is adequate for most financial returns. The GARCH(1,1)-model can be written as:

$$\sigma_t^2 = \gamma V_L + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2 \, , \tag{2.2.2.3}$$

where $\gamma + \alpha + \beta = 1$, and V_L is the long-run average variance rate. Defining

$$\omega = \gamma V_L$$

the model (2.2.2.3) can be written as

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2. \tag{2.2.2.4}$$

Maximum likelihood methods can be used to obtain estimates for parameters in (2.2.2.4). Assuming that r_t, conditional on the volatility, is normally distributed (most studies have shown this to be a good approximation), and that we have n observations, we would then want to find the parameters ω, α and β that maximize:

$$\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma_i}\exp(\frac{-\varepsilon_i^2}{2\sigma_i^2})\right)$$

or

$$\sum_{i=1}^{n}(-\log(\sigma_i^2) - \frac{\varepsilon_i^2}{\sigma_i^2}) .$$

An estimate for the long-run volatility can then be found by $\widehat{V_L} = \frac{\widehat{\omega}}{1-\widehat{\alpha}-\widehat{\beta}}$.

Theorem 2.1.1

After the parameters have been estimated, forecasts of volatility can be found by

$$\widehat{\sigma}_{t+s}^2 = \widehat{V}_L + (\widehat{\alpha} + \widehat{\beta})^s(\widehat{\sigma}_t^2 - \widehat{V_L}) .$$

Proof

From (2.2.2.4)

$$\begin{aligned}\sigma_t^2 &= \omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2 \\ &= (1-\alpha-\beta)V_L + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2\end{aligned}$$

$$\Rightarrow \sigma_t^2 - V_L = \beta(\sigma_{t-1}^2 - V_L) + \alpha(\varepsilon_{t-1}^2 - V_L) .$$

At time $t+s$ in the future,

$$\sigma_{t+s}^2 - V_L = \beta(\sigma_{t+s-1}^2 - V_L) + \alpha(\varepsilon_{t+s-1}^2 - V_L)$$

$$\Rightarrow E(\sigma_{t+s}^2 - V_L) = (\alpha+\beta)E(\sigma_{t+s-1}^2 - V_L) \text{ because } E(\varepsilon_{t+s-1}^2) = E(\sigma_{t+s-1}^2)\,.$$

Remembering that $E\left(\sigma_{t+1}^2\right) = \sigma_t^2$, then it follows by recursion that,

$$E(\sigma_{t+s}^2 - V_L) = (\alpha+\beta)^s(\sigma_t^2 - V_L)$$

or

$$E(\sigma_{t+s}^2) = V_L + (\alpha+\beta)^s(\sigma_t^2 - V_L)$$

which completes the proof.

The GARCH-model in the form of (2.2.2.1) does not take the leverage effect into account. A volatility model that does take the leverage effect into account is the so-called EGARCH-model.

2.2.3 The EGARCH- model

Nelson (1991) suggested the following volatility model:

$$\log(\sigma_t^2) = \omega + \beta\log(\sigma_{t-1}^2) + \lambda\frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \phi(\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| - \left(\frac{2}{\pi}\right)^{\frac{1}{2}})\,,$$

with $\omega > 0, \phi$ is a unknown constant and λ measures the asymmetric effect that the residual returns have on the variance. Consequently if $\lambda < 0$, negative past residual returns will have a greater influence on the conditional (on past information) variance $\log(\sigma_t^2)$ than positive residual returns have.

2.3 Diagnostic tests for ARCH-type models

After some or other model has been decided on, tests need to be carried out to determine whether the chosen model is rejected or not. Most of the tests dealing with misspecification, inspect the properties of the standardized residual returns, defined as

$$\varepsilon_t^* = \frac{\varepsilon_t}{\sigma_t}\,,$$

where σ_t^2 is the variance that is assumed by the model in question.

If the model is correct, then ε_t^* should be white noise. Suppose we had a sample of ε_i^* , $i = 1, 2, ...n$, then one way of testing would be to examine the autocorrelation of $(\varepsilon_i^*)^2$, $i = 1, 2, ...n$. For a more scientific test, the Ljung-Box statistic may be used. The Ljung-Box statistic is

$$n \sum_{j=1}^{T} \nu_j \eta_j^2 ,$$

where η_j is the autocorrelation of $(\varepsilon_i^*)^2$, $i = 1, 2, ...n$ of lag j, and $\nu_j = \frac{m-2}{m-j}$.

Critical values for the Ljung-Box statistic can be found in literature [See Hull, 2003].

2.4 Problems during stage two

Since volatility is inherently unobservable the techniques developed during the second stage had been learnt by studying volatility implied by options or by studying direct indicators of volatility. However all these techniques have weaknesses. To illustrate: implied volatility is based on some or other model (such as Black Scholes), but these models have the risk of model misspecification. Direct indicators such as ex-post (historical) squared returns are contaminated by noise. Anderson and Bollershev(1998) documented that the variance of the noise is typically too large relative to that of the signal. Therefore, neither ex- post squared returns nor implied volatility is a good volatility measurement.

All the diagnostic tests for a specific model reject only wrong models with a certain probability, but cannot indicate whether the model is correct. Therefore, a reliable model could only be built if all the stylised facts of volatility were known. **As has been pointed out above, only a model-free and error-free volatility measurement can catch all the stylised facts**. Another reason why a good volatility measurement was needed is that if a model is correct, then the better the measurements are, the better the forecasts will be. Also, forecasting models (ex-ante expected volatility) can then be tested more efficiently.

Researchers realised the importance of measuring ex-post volatility first, and then concentrating on ex-ante expected volatility. When the new measurement (discussed in Chapter 4) of volatility was employed, the following stylised facts of volatility and returns were found.

Stylised facts of returns and volatility:

i) Fat tails of high frequency returns (kurtosis >3), and the kurtosis tends to decrease as the time interval increases. This is called the cross-over effect.

ii) High frequency returns are autocorrelated.

iii) The volatility measurements are autocorrelated for up to at least a month (volatility clustering).

iv) Returns have the multifractal property: $\mathrm{E}(|p_{t+\Delta t} - p_t|^q) \sim (\Delta t)^{q \cdot H(q)}$ where $H\,(q)$ is the Hurst exponent.

v) Volatiliy cascade characteristic: long-term views have a marked influence on the short-term views. Volatility over longer intervals has a larger influence on shorter intervals than conversely. (This is more important economically than mathematically.) Volatility over shorter intervals can be written as a function of volatility over longer time intervals.

vi) Leverage effect: volatility tends to increase when prices are falling, and vice versa.

vii) $\frac{r_t}{\sigma_t}$ is normally distributed.

viii) σ_t is lognormally distributed.

The above mentioned models have only a few of these desired characteristics. The GARCH model only caters for the fat tail of returns and volatility clustering while the EGARCH model captures only the fat tails, volatility clustering and the leverage effect of volatility. It is thus clear that a new volatility measurement was needed.

3

Estimation for discretely observed diffusions

3.1 Introduction

In this chapter a parametric technique for measuring volatility, if the time horizon tends to zero, is discussed. The volatility, if the time horizon tends to zero, is called instantaneous volatility. This particular technique is published in Kelly, Platen and Sϕrensen (2003). [See Dorogovcev, 1976, Hansen, 1982, Prakasa Rao,1988 and Kessler, 1997, amongst other, where other techniques are discussed.] In Chapter 4 we return to the general case of larger time horizons, but here the estimation of the parameters of a stochastic differential equation (SDE) for discretely observed diffusions is discussed. The theory behind an SDE based on stock returns, needs to be dealt with first.

It is generally assumed that asset returns can be decomposed into two parts: a predictable part and an unpredictable part. The predictable part is perceived as a very simple process, while the unpredictable part will be modelled with a martingale process, discussed in Section 3.2. The two parts can be seen as a drift part and a diffusion part, respectively, with each part consisting of an arbitrary number of parameters. The diffusion part is the instantaneous volatility to be estimated. In Section 3.3 a parameter estimation technique for the two parts is given. This method can be used in non-financial situations as well.

3.2 Martingale theory and Ito's integral

The conditional distribution of a random variable X_t given $\boldsymbol{\theta} : p \times 1$ is denoted by $X_t \mid \boldsymbol{\theta}$. Here $\boldsymbol{\theta}$ may be a set of random variables or a set of parameters. Let F_t be the information set or set of events at time t, with $F_s \subset F_t$, $s \leq t$ and $t \in [0, T]$. F_t is referred to as a σ-field. Further, let P denote a probability measure defined on (Ω, P, F_t), where Ω denotes the sample space (states of the world), and F_t the set of events up to time t. Let M_t be a random process, with $t \in [0, T]$. If the value of M_t is known, given the information set F_t, the process $\{M_t, t \in [0, T]\}$ is said to be adapted to $\{F_t, t \in [0, T]\}$.

Definition 3.2.1 (Neftci, 2000)

The process $\{M_t, t \in [0, T]\}$ is a martingale with respect to F_t and P, if, for all $t \geq 0$:

i) M_t is F_t-adapted,

ii) $\mathrm{E}(\mid M_t \mid) < \infty$ and

iii) $\mathrm{E}(M_t \mid F_s) = M_s$, for all $s \leq t, t \in [0, T]$.

Furthermore, if $\mathrm{E}(M_t^2) < \infty$, then M_t is said to be square integrable. In martingale theory the values of the stochastic process $\{M_t, t \in [0, T]\}$ is often needed at particular points in the time interval $[0, T]$. This is done by partitioning the interval $[0, T]$ as follows:

$$0 = t_0 < t_1 < \ ... \ < t_{n-1} < t_n = T \ .$$

We assume the partioning is such that as $n \to \infty$, the interval $[0, T]$ is partitioned into finer and finer intervals.

Definition 3.2.2 (Neftci, 2000)

A Wiener process W_t, relative to a family of information sets $\{F_t\}$, is a stochastic process such that :

i) W_t is a square integrable martingale with $W_o = 0$, and $\mathrm{E}\left[(W_t - W_s)^2\right] = t - s$, where $s \leq t$.

ii) The trajectories of W_t are continuous over t.

iii) Finally, the process is continuous, i.e. in infinitesimal intervals, the movements of W_t are infinitesimal.

This definition implies the following properties of a Wiener process:

- W_t has uncorrelated increments because it is a martingale, and because every martingale has unpredictable increments.
- W_t has zero mean because it starts at zero, and the mean of every increment equals zero.
- W_t has variance t.
- $W_t - W_s \sim N(0, |t-s|)$

Definition 3.2.3 (Neftci, 2000)

$\int_0^t M_{s-}dM_s$ with M_s a stochastic variable, and $M_{s-} = \lim_{v\uparrow s, v\leq s} M_v$ is called an Ito integral if

i) $\mathrm{E}\left(\int_0^t M_s^2 ds\right) < \infty$

ii) $\lim_{n\to\infty}\mathrm{E}\left[\sum_{k=1}^n M_{t_{k-1}}\left(M_{t_k} - M_{t_{k-1}}\right) - \int_0^t M_{s-}dM_s\right]^2 = 0$ (Mean square)

A return model can now be constructed that consists of a predictable and an unpredictable part.

Definition 3.2.4 Ito's formula (Neftci, 2000)

Let S_t be the market value of the share or index at time t. Let $F\left(S_t, t\right)$ be a twice differentiable function of t and of the random process S_t, where

$$dS_t = \mu\left(t, S_t\right)dt + \sigma\left(t, S_t\right)dW_t\,, \tag{3.2.1}$$

with W_t a Wiener process with respect to some information set. The term $\mu\left(t, S_t\right)dt$ denotes the predictable process , and $\sigma\left(t, S_t\right)dW_t$ denotes the unpredictable part. It then follows that

$$dF_t = \frac{\partial F}{\partial S_t}dS_t + \frac{\partial F}{\partial t}dt + \frac{1}{2}\frac{\partial^2 F}{\partial S_t^2}\sigma^2\left(t, S_t\right)dt \tag{3.2.2}$$

or, after substituting (3.2.1) in (3.2.2),

$$dF_t = \left[\mu\left(t, S_t\right)\frac{\partial F}{\partial S_t} + \frac{\partial F}{\partial t} + \sigma^2\left(t, S_t\right)\frac{1}{2}\frac{\partial^2 F}{\partial S_t^2}\right]dt + \sigma^2\left(t, S_t\right)\frac{\partial F}{\partial S_t}dW_t\,,$$

where the equality holds in the mean square sense.

3.3 Parameter estimation

Consider the following SDE :

$$dS_t = \mu(t, S_t, \boldsymbol{\theta})\, dt + \sigma(t, S_t, \boldsymbol{\theta})\, dW_t\,, \tag{3.3.1}$$

where W_t is a Wiener process with respect to some information set , and $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)^T$.

We want to estimate the parameter values $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)^T$. We assume that the SDE has a unique solution for all parameter values $\boldsymbol{\theta}$ in a given open subset $\Theta \subseteq \mathbb{R}^p \in \{1, 2, ...\}$. The drift and diffusion coefficient functions $\mu(t, S_t, \boldsymbol{\theta})$ and $\sigma(t, S_t, \boldsymbol{\theta})$ respectively in (3.3.1), are assumed to be known, with the exception of the parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)^T \in \Theta$. The functional form of $\mu(t, S_t, \boldsymbol{\theta})$ and $\sigma(t, S_t, \boldsymbol{\theta})$ is for the individual to decide upon. Later this technique will be illustrated using as drift and coefficient functions:

$$\mu(t, S_t, \boldsymbol{\theta}) \equiv (\theta_1 + \theta_2 S_t) \qquad \text{and} \qquad \sigma(t, S_t, \boldsymbol{\theta}) \equiv \sqrt{\theta_3 + \theta_4 S_t}\,.$$

Partition the interval $[0, T]$ into n subintervals of equal lengths $\Delta = t_k - t_{k-1}$ for $k = 1, 2, ..., n$, with $t_0 = 0, t_n = T$. Assume we observe at each t_k for $k = 0, 1, ..., n$, the prices S_{t_k}. Let $F(t_k, S_{t_k}, \lambda_i)$ be a function of S_{t_k}, where λ_i for $i = 1, 2, ..., p$ are the transformed parameters and can be chosen freely. The function $F(t_k, S_{t_k}, \lambda_i)$ is called the i-th transform function, and is used to transform the prices in a manner that allows us to obtain good estimates of the unknown parameters. $F(t_k, S_{t_k}, \lambda_i)$ will be taken as $S_t^{\lambda_i}$. This form of $F(t_k, S_{t_k}, \lambda_i)$ is found to give good estimates of the unknown parameters. [See Kelly, Platen and Sϕrensen 2000.]

The parameter estimation technique published in Kelly, Platen and Sϕrensen (2003) consists of obtaining a set of linear equations via the use of transform functions ($F(t_k, S_{t_k}, \lambda_i)$), with the only unknowns the elements in $\boldsymbol{\theta}$. We will show that the closer the observation times are, the closer the expected value of the particular set of equations will be to zero. The derivation of the equations is mathematically intensive and the reason for taking this route will become clearer only after the derivation of the equations. Some lemmas and theorems need to be given first before the set of equations can be derived.

Applying Ito's formula to the transformed prices leads to

$$
\begin{aligned}
dF\left(t, S_t, \lambda_i\right) &= \underbrace{\left[\mu\left(t, S_t, \boldsymbol{\theta}\right)\frac{\partial F}{\partial S_t} + \frac{\partial F}{\partial t} + \sigma^2\left(t, S_t, \boldsymbol{\theta}\right)\frac{1}{2}\frac{\partial^2 F}{\partial S_t^2}\right]}_{=L_0[F(t,S_t,\lambda_i)]} dt + \\
& \underbrace{\sigma\left(t, S_t, \boldsymbol{\theta}\right)\frac{\partial F}{\partial S_t}}_{=L_1[F(t,S_t,\lambda_i)]} dW_t \qquad (3.3.2) \\
\text{for } i &= 1, 2, ..., p.
\end{aligned}
$$

For $i = 1, 2, ..., p$ and $k = 0, 1, ..., n$ define the quantities,

$$D_{\lambda_i,k,\Delta} = \frac{1}{t_k - t_{k-1}}\left(F(t_k, S_{t_k}, \lambda_i) - F(t_{k-1}, S_{t_{k-1}}, \lambda_i)\right) \qquad (3.3.3)$$

and

$$Q_{\lambda_i,k,\Delta} = \frac{1}{t_k - t_{k-1}}(F(t_k, S_{t_k}, \lambda_i) - F(t_{k-1}, S_{t_{k-1}}, \lambda_i))^2\,. \qquad (3.3.4)$$

The definitions in (3.3.3) and (3.3.4) will soon become clear. In the following lemma an important result that is used in the estimation of the parameters is proved.

Lemma 3.3.1

$$E\left(D_{\lambda_i,k,\Delta} \mid S_{t_{k-1}}\right) = L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)] + o\left(\Delta\right). \qquad (3.3.5)$$

Proof

We have from (3.3.2) that

$$dF\left(t, S_t, \lambda_i\right) = L_0[F\left(t, S_t, \lambda_i\right)]dt + L_1[F\left(t, S_t, \lambda_i\right)]dW_t$$

and hence

$$
\begin{aligned}
\int_{t_{k-1}}^{t_k} dF\left(t, S_t, \lambda_i\right) &= F(t_k, S_{t_k}, \lambda_i) - F(t_{k-1}, S_{t_{k-1}}, \lambda_i) \\
&= \int_{t_{k-1}}^{t_k} L_0[F\left(t, S_t, \lambda_i\right)]dt + \int_{t_{k-1}}^{t_k} L_1[F\left(t, S_t, \lambda_i\right)]dW_t\,. \qquad (3.3.6)
\end{aligned}
$$

First consider the second term in (3.3.6)

$$\begin{aligned}
\mathbf{I}_1 &= \int_{t_{k-1}}^{t_k} L_1[F(t,S_t,\lambda_i)]dW_t \\
&= \int_{t_{k-1}}^{t_k}\int_0^t dL_1[F(u,S_u,\lambda_i)]dW_t \text{ because } L_1[F(0,S_0,\lambda_i)]=0 \\
&= \int_{t_{k-1}}^{t_k}(\int_0^{t_{k-1}} dL_1[F(u,S_u,\lambda_i)] + \int_{t_{k-1}}^{t} dL_1[F(u,S_u,\lambda_i)])dW_t\,. \qquad (3.3.7)
\end{aligned}$$

Applying Ito's formula to the second term in (3.3.7), leads to

$$\begin{aligned}
\mathbf{I}_1 &= \int_{t_{k-1}}^{t_k} L_1[F(t_{k-1},S_{t_{k-1}},\lambda_i)]dW_t + \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t}(L_0[L_1[F(u,S_u,\lambda_i)]]du \\
&\quad + L_1[L_1[F(u,S_u,\lambda_i)]]dW_u)dW_t \\
&= L_1[F(t_{k-1},S_{t_{k-1}},\lambda_i)]\int_{t_{k-1}}^{t_k} dW_t + \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t} L_0[L_1[F(u,S_u,\lambda_i)]]dudW_t \\
&\quad + \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t} L_1[L_1[F(u,S_u,\lambda_i)]]dW_u dW_t \\
&= \underbrace{L_1[F(t_{k-1},S_{t_{k-1}},\lambda_i)]\left[W_{t_k}-W_{t_{k-1}}\right]}_{(1)} + \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t}\int_0^u dL_0[L_1[F(w,S_w,\lambda_i)]]dudW_t \\
&\quad + \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t}\int_0^u dL_1[L_1[F(w,S_w,\lambda_i)]]dW_u dW_t \\
&= (1) + \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t}\int_0^{t_{k-1}} dL_0[L_1[F(w,S_w,\lambda_i)]]dudW_t \\
&\quad \underbrace{+ \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t}\int_{t_{k-1}}^{u} dL_0[L_1[F(w,S_w,\lambda_i)]]dudW_t}_{remainder\ term}
\end{aligned}$$

$$
\begin{aligned}
& + \underbrace{\int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} \int_{0}^{t_{k-1}} dL_1[L_1[F(w, S_w, \lambda_i)]]dW_u dW_t + \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} \int_{t_{k-1}}^{u} dL_1[L_1[F(w, S_w, \lambda_i)]]dW_u dW_t}_{remainder\ term} \\
= \;& (1) + \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} L_0[L_1[F(t_{k-1}, S_{t_{k-1}}, \lambda_i)]]du dW_t \\
& + \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} L_1[L_1[F(t_{k-1}, S_{t_{k-1}}, \lambda_i)]]dW_u dW_t + remainder\ terms \\
= \;& (1) + \underbrace{L_0[L_1[F(t_{k-1}, S_{t_{k-1}}, \lambda_i)]] \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} du dW_t}_{(2)} \\
& + L_1[L_1[F(t_{k-1}, S_{t_{k-1}}, \lambda_i)]] \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} dW_u dW_t + remainder\ terms \\
= \;& (1) + (2) + L_1[L_1[F(t_{k-1}, S_{t_{k-1}}, \lambda_i)]] \int_{t_{k-1}}^{t_k} \left(W_t - W_{t_{k-1}}\right) dW_t + remainder\ terms \\
= \;& (1) + (2) + L_1[L_1[F(t_{k-1}, S_{t_{k-1}}, \lambda_i)]] \cdot \\
& \left[\int_{t_{k-1}}^{t_k} W_t dW_t - W_{t_{k-1}} \left(W_{t_k} - W_{t_{k-1}}\right)\right] + remainder\ terms \\
= \;& (1) + (2) + L_1[L_1[F(t_{k-1}, S_{t_{k-1}}, \lambda_i)]] \cdot \\
& \left[\int_{0}^{t_k} W_t dW_t - \int_{0}^{t_{k-1}} W_t dW_t - W_{t_{k-1}} \left(W_{t_k} - W_{t_{k-1}}\right)\right] + remainder\ terms\,.
\end{aligned}
$$

To solve

$$\int_{t_{k-1}}^{t_k} W_t dW_t$$

choose

$$F(W_t, t) = \frac{1}{2} W_t^2\,.$$

The choice of $\frac{1}{2}W_t^2$ for $F(W_t, t)$ follows from the following manipulations. Applying Ito's formula to $F\left(W_t, t\right)$ leads to

$$
\begin{aligned}
dF\left(W_t, t\right) &= 0 + W_t dW_t + \frac{1}{2} dt\,. \\
&\Rightarrow \; F\left(W_t, t\right) = \int_0^t W_s dW_s + \frac{1}{2} t \\
&\Rightarrow \; \frac{1}{2} W_t^2 = \int_0^t W_s dW_s + \frac{1}{2} t \\
&\Rightarrow \; \int_0^t W_s dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} t.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& \mathbf{I}_1 = \int_{t_{k-1}}^{t_k} L_1[F\left(t, S_t, \lambda_i\right)] dW_t \\
= \;& (1) + (2) + L_1[L_1[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]] \cdot \\
& \left[\frac{1}{2} W_{t_k}^2 - \frac{1}{2} t_k - \left(\frac{1}{2} W_{t_{k-1}}^2 - \frac{1}{2} t_{k-1}\right) - W_{t_{k-1}} W_{t_k} + W_{t_{k-1}}^2\right] + remainder\ terms \\
= \;& (1) + (2) + L_1[L_1[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]] \cdot \\
& \frac{1}{2}\left[W_{t_k}^2 - 2W_{t_{k-1}} W_{t_k} + W_{t_{k-1}}^2 - \left(t_k - t_{k-1}\right)\right] + remainder\ terms \\
= \;& L_1[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]\left[W_{t_k} - W_{t_{k-1}}\right] + \\
& L_0 L_1 F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right) \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} dt dW_t \\
& + L_1[L_1[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]] \frac{1}{2}\left[\left(W_{t_k} - W_{t_{k-1}}\right)^2 - \left(t_k - t_{k-1}\right)\right] \\
& + remainder\ terms\,. \qquad (3.3.8)
\end{aligned}
$$

Now consider the first term in (3.3.6).

$$
\begin{aligned}
\mathbf{I}_2 &= \int_{t_{k-1}}^{t_k} L_0[F\left(t, S_t, \lambda_i\right)] dt \\
&= \int_{t_{k-1}}^{t_k} \int_0^t dL_0[F\left(u, S_u, \lambda_i\right)] dt \text{ because } L_0 F\left(0, S_0, \lambda_i\right) = 0 \\
&= \int_{t_{k-1}}^{t_k} \int_0^{t_{k-1}} dL_0[F\left(u, S_u, \lambda_i\right)] dt + \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} dL_0[F\left(u, S_u, \lambda_i\right)] dt. \qquad (3.3.9)
\end{aligned}
$$

Applying Ito's formula to the second term in (3.3.9) leads to

$$
\begin{aligned}
\mathbf{I}_2 &= \underbrace{L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]\left(t_k - t_{k-1}\right)}_{(3)} \\
&\quad + \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} \left(L_0[L_0[F\left(u, S_u, \lambda_i\right)]]du + L_1[L_0[F\left(u, S_u, \lambda_i\right)]]dW_u\right) dt \\
&= (3) + \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} L_0[L_0[F\left(u, S_u, \lambda_i\right)]]dudt \\
&\quad + \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} L_1[L_0[F\left(u, S_u, \lambda_i\right)]]dW_u dt \\
&= (3) + \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} \int_0^u dL_0[L_0[F\left(w, S_w, \lambda_i\right)]]dudt \\
&\quad + \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} \int_0^u dL_1[L_0[F\left(w, S_w, \lambda_i\right)]]dudt \\
&= (3) + \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} \int_0^{t_{k-1}} dL_0[L_0[F\left(w, S_w, \lambda_i\right)]]dudt \\
&\quad \underbrace{+ \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} \int_{t_{k-1}}^{t} dL_0[L_0[F\left(w, S_w, \lambda_i\right)]]dudt}_{remainder\ term} \\
&\quad + \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} \int_0^{t_{k-1}} dL_1[L_0[F\left(w, S_w, \lambda_i\right)]]dW_u dt \\
&\quad \underbrace{+ \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} \int_{t_{k-1}}^{t} dL_1[L_0[F\left(w, S_w, \lambda_i\right)]]dW_u dt}_{remainder\ term} \\
&= (3) + L_0[L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]] \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} dudt \\
&\quad \underbrace{+ L_1[L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]] \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^{t} dW_u dt}_{(4)} + remainder\ terms \\
&= (3) + (4) + L_0[L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]] \cdot \\
&\quad \int_{t_{k-1}}^{t_k} \left(t - t_{k-1}\right) dt + remainder\ terms \\
&= (3) + (4) + L_0[L_0[F t_{k-1}, S_{t_{k-1}}, \lambda_i)]] \cdot \\
&\quad \left(\frac{1}{2}\left(t_k^2 - t_{k-1}^2\right) - t_{k-1}\left(t_k - t_{k-1}\right)\right) + remainder\ terms
\end{aligned}
$$

$$
\begin{aligned}
&= (3)+(4)+L_0[L_0]Ft_{k-1}, S_{t_{k-1}}, \lambda_i)]] \cdot \\
&\quad \frac{1}{2}\left(t_k^2 + t_{k-1}^2 - 2t_k t_{k-1}\right) + remainder\ terms \\
&= (3)+(4)+L_0[L_0[Ft_{k-1}, S_{t_{k-1}}, \lambda_i)]] \cdot \\
&\quad \frac{1}{2}\left(t_k - t_{k-1}\right)^2 + remainder\ terms \\
&= L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]\left(t_k - t_{k-1}\right) + \\
&\quad L_1[L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]] \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t} dW_u dt + \\
&\quad L_0[L_0[F(t_{k-1}, S_{t_{k-1}}, \lambda_i)]]\frac{1}{2}\left(t_k - t_{k-1}\right)^2 + remainderterms\ . \qquad (3.3.10)
\end{aligned}
$$

By substituting (3.3.9) and (3.3.10) in the second and first term of (3.3.6) respectively, it follows that

$$
\begin{aligned}
& F(t_k, S_{t_k}, \lambda_i) - F(t_{k-1}, S_{t_{k-1}}, \lambda_i) \\
=\ & L_1[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]\left[W_{t_k} - W_{t_{k-1}}\right] + L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]\left(t_k - t_{k-1}\right) \\
& + L_1[L_1[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]]\frac{1}{2}\left[(W_{t_k} - W_{t_{k-1}})^2 - (t_k - t_{k-1})\right] + \\
& L_0[L_0[F(t_{k-1}, S_{t_{k-1}}, \lambda_i)]]\frac{1}{2}\left(t_k - t_{k-1}\right)^2 \\
& + L_0[L_1[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]] \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t} dt dW_t + \\
& L_1[L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]] \int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t} dW_u dt + remainder\ terms\ ,
\end{aligned}
$$

since $E\left[W_{t_k} - W_{t_{k-1}}\right] = 0$, $E\left[W_{t_k} - W_{t_{k-1}}\right]^2 = \Delta$ and the remainder terms are of order Δ^3. Further

$$
\begin{aligned}
E\left[\int_{t_{k-1}}^{t_k}\int_{t_{k-1}}^{t} dt dW_t\right] &= E\left[\int_{t_{k-1}}^{t_k}\left(t - t_{k-1}\right) dW_t\right] \\
&= \Delta E\left[\int_{t_{k-1}}^{t_k} dW_t\right] = \Delta E\left(W_{t_k} - W_{t_{k-1}}\right) \\
&= 0.
\end{aligned}
$$

Hence

$$
\begin{aligned}
E\left(D_{\lambda_i,k,\Delta} \mid S_{t_{k-1}}\right) &= 0 + L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)] + 0 + \frac{1}{2}L_0\left[L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]\right]\Delta \\
&\quad +0+0+o\left(\Delta^2\right) \\
&= L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)] + o\left(\Delta\right) \ ,
\end{aligned}
$$

which proves the lemma.

The set of equations will now be defined and thereafter it will be shown, using Lemma 3.3.1, that the closer the observation times are, the closer the expected value of the set of equations will be to zero. In Section 3.4 an example is given.

Define

$$
\mathbf{F}_k\left(\boldsymbol{\theta}\right) : 2p \times 1 = \left(\mathbf{F}_k^{(1)}\left(\boldsymbol{\theta}\right)^T, \mathbf{F}_k^{(2)}\left(\boldsymbol{\theta}\right)^T\right)^T \ ,
$$

with

$$
\mathbf{F}_k^{(j)}\left(\boldsymbol{\theta}\right) : p \times 1 = \left(F_{k,1}^{(j)}\left(\boldsymbol{\theta}\right), F_{k,2}^{(j)}\left(\boldsymbol{\theta}\right), ..., F_{k,p}^{(j)}\left(\boldsymbol{\theta}\right)\right)^T, j \in \{1,2\} \ ,
$$

where

$$
\begin{aligned}
F_{k,i}^{(1)}\left(\boldsymbol{\theta}\right) &= D_{\lambda_i,k,\Delta} - L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)], i \in \{1,2,...,p\} \\
&= \frac{1}{t_k - t_{k-1}}\left(F\left(t_k, S_{t_k}, \lambda_i\right) - F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)\right) - L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)] \\
i &\in \{1,2,...,p\}
\end{aligned}
$$

and

$$
\begin{aligned}
F_{k,i}^{(2)}\left(\boldsymbol{\theta}\right) &= Q_{\lambda_i,k,\Delta} - \left(L_1[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]\right)^2, i \in \{1,2,...,p\} \\
&= \frac{1}{t_k - t_{k-1}}\left(F\left(t_k, S_{t_k}, \lambda_i\right) - F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)\right)^2 - L_1[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)] \\
i &\in \{1,2,...,p\} \ .
\end{aligned}
$$

The class of estimating functions given in Kelly, Platen and Sϕrensen (2003) is

$$
\mathbf{K}\left(\boldsymbol{\theta}, t, \Delta\right) : p \times 1 = \frac{1}{n}\sum_{k=1}^{n}\mathbf{M}\left(\boldsymbol{\theta}\right)\mathbf{F}_k\left(\boldsymbol{\theta}\right) \ ,
$$

where

$$\mathbf{M}(\boldsymbol{\theta}) : p \times 2p = \mathbf{M}\left(\boldsymbol{\theta}, t_{k-1}, S_{t_{k-1}}, \Delta\right)$$

is free to be chosen appropriately.

Theorem 3.3.1

The elements in

$$E\left(\mathbf{K}(\boldsymbol{\theta}, t, \Delta) \mid S_{t_{k-1}}\right) \text{ are of order } \Delta.$$

Proof

$$\begin{aligned} E\left(\mathbf{K}(\boldsymbol{\theta}, t, \Delta) \mid S_{t_{k-1}}\right) &= E\left(\frac{1}{n}\sum_{k=1}^{n}\mathbf{M}(\boldsymbol{\theta})\,\mathbf{F}_k(\boldsymbol{\theta}) \mid S_{t_{k-1}}\right) \\ &= \frac{1}{n}\sum_{k=1}^{n}\mathbf{M}(\boldsymbol{\theta})\,E(\mathbf{F}_k(\boldsymbol{\theta}) \mid S_{t_{k-1}}) \end{aligned}$$

because $M(\theta)$ depends only on $S_{t_{k-1}}$. Hence consider

$$\begin{aligned} E(F_{k,i}^{(1)}(\boldsymbol{\theta}) \quad \mid \quad & S_{t_{k-1}}) = E(D_{\lambda_i,k,\Delta} - L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)] \mid S_{t_{k-1}}) \\ = \quad & E(D_{\lambda_i,k,\Delta} \mid S_{t_{k-1}}) - L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)] \\ = \quad & L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)] + o(\Delta) - L_0[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)] \\ = \quad & o(\Delta)\ . \end{aligned}$$

Next consider

$$\begin{aligned} E(F_{k,i}^{(2)}(\boldsymbol{\theta}) \quad \mid \quad & S_{t_{k-1}}) \\ = \quad & E\left(Q_{\lambda_i,k,\Delta} - \left(L_1 F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)\right)^2 \mid S_{t_{k-1}}\right) \\ = \quad & E(\frac{1}{t_k - t_{k-1}}(F(t_k, S_{t_k}, \lambda_i) - F(t_{k-1}, S_{t_{k-1}}, \lambda_i))^2 \\ & - \left(L_1[F\left(t_{k-1}, S_{t_{k-1}}, \lambda_i\right)]\right)^2 | S_{t_{k-1}}) \end{aligned}$$

From the middle of page 29 it follows that

$$\begin{aligned}
E(F_{k,i}^{(2)}(\boldsymbol{\theta}) &\mid S_{t_{k-1}}) \\
&= \frac{1}{t_k - t_{k-1}} E\left((L_1 F(t_{k-1}, S_{t_{k-1}}, \lambda_i))^2 \left[W_{t_k} - W_{t_{k-1}} \right]^2 \mid S_{t_{k-1}} \right) \\
&\quad - \left(L_1 F(t_{k-1}, S_{t_{k-1}}, \lambda_i) \right)^2 + o\left(\Delta^2\right) \\
&= L_1 F(t_{k-1}, S_{t_{k-1}}, \lambda_i))^2 \frac{1}{t_k - t_{k-1}} E(W_{t_k} - W_{t_{k-1}})^2 \\
&\quad - \left(L_1 F(t_{k-1}, S_{t_{k-1}}, \lambda_i) \right)^2 + o\left(\Delta^2\right) \\
&= L_1 F(t_{k-1}, S_{t_{k-1}}, \lambda_i))^2 \frac{1}{t_k - t_{k-1}} (t_k - t_{k-1}) \\
&\quad - \left(L_1 F(t_{k-1}, S_{t_{k-1}}, \lambda_i) \right)^2 + o\left(\Delta^2\right) \\
&= o\left(\Delta^2\right) ,
\end{aligned}$$

which completes the proof.

It follows that the closer the observation times are, the closer the expected value of

$$\mathbf{K}(\boldsymbol{\theta}, t, \Delta) \mid S_{t_{k-1}}$$

will be to zero. Thus by setting

$$\frac{1}{n} \sum_{k=1}^{n} \mathbf{M}(\boldsymbol{\theta}) \mathbf{F}_k(\boldsymbol{\theta}) = \mathbf{0}$$

and choosing values for λ_i, $i = \{1, 2, ...p\}$, we get a system of p equations with the only unknown values being the parameters in $\boldsymbol{\theta}$.The exact choice for the values of λ_i, $i = \{1, 2, ...p\}$ is not important as different choices should lead to more or less the same results. By solving this system, estimates for the parameters in $\boldsymbol{\theta}$ are obtained.

The estimation procedure is simplified when the parameter vector $\boldsymbol{\theta}$ can be written as $\boldsymbol{\theta} : p \times 1 = \left(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T \right)^T$, where $\boldsymbol{\alpha} : p_1 \times 1$ appears only in the drift coefficient , and $\boldsymbol{\beta} : p_2 \times 1$ appears only in the diffusion coefficient. In this case we first estimate $\boldsymbol{\beta}$ by solving

$$\mathbf{H}(\boldsymbol{\beta}, t, \Delta) : p \times 1 = \frac{1}{n} \sum_{k=1}^{n} \mathbf{M}_1(\boldsymbol{\theta}) \mathbf{F}_k^{(2)}(\boldsymbol{\beta}) = \mathbf{0} \tag{3.3.11}$$

and then by solving

$$\mathbf{G}\left(\boldsymbol{\alpha},\boldsymbol{\beta},t,\Delta\right):p\times 1=\frac{1}{n}\sum_{k=1}^{n}\mathbf{M}_{2}\left(\boldsymbol{\theta}\right)\mathbf{F}_{k}^{(1)}\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right)=\mathbf{0} \tag{3.3.12}$$

α is estimated. By solving (3.3.12), $\boldsymbol{\beta}$ is replaced by its estimate acquired when (3.3.11) was solved.

For a more in-depth analysis of this estimation method see Kelly, Platen and Sϕrensen (2003). In the following example the simplified procedure discussed above is illustrated.

3.4 An example

Let the shifted square root process be given by

$$dS_t=(\theta_1+\theta_2 S_t)dt+\sqrt{\theta_3+\theta_4 S_t}dW_t\ .$$

With the above notation it follows that

$$\mu\left(t,S_t\right)\equiv\left(\theta_1+\theta_2 S_t\right)$$

and

$$\sigma\left(t,S_t\right)\equiv\sqrt{\theta_3+\theta_4 S_t}\ .$$

Let the transform function be

$$F\left(t,S_t,\lambda_i\right)=S_t^{\lambda_i}$$

and choose

$$\mathbf{M}_1\left(\boldsymbol{\theta}\right)=I_{2\times 2}=\mathbf{M}_2\left(\boldsymbol{\theta}\right)=\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

so that

$$H_i\left(\boldsymbol{\beta},t,\Delta\right)=\frac{1}{n}\sum_{k=1}^{n}F_{k,i}^{(2)}\left(\boldsymbol{\beta}\right)=0, for\ i\in\{3,4\}\ ,$$

with

$$F_{k,i}^{(2)}\left(\boldsymbol{\beta}\right)=Q_{\lambda_i,k,\Delta}-\left(L_1F\left(t_{k-1},S_{t_{k-1}},\lambda_i\right)\right)^2$$

and

$$G_i\left(\boldsymbol{\alpha},\boldsymbol{\beta},t,\Delta\right)=\frac{1}{n}\sum_{k=1}^{n}F_{k,i}^{(1)}\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right)=0,\ \text{for } i\in\{1,2\},$$

with

$$F_{k,i}^{(1)}\left(\beta\right)=D_{\lambda_i,k,\Delta}-L_0F\left(t_{k-1},S_{t_{k-1}},\lambda_i\right)\ .$$

Hence

$$\begin{aligned} L_0[F\left(t_{k-1},S_{t_{k-1}},\lambda_i\right)] &= \mu\left(t,S_t,\boldsymbol{\alpha}\right)\frac{\partial F}{\partial S_t}+\frac{\partial F}{\partial t}+\sigma^2\left(t,S_t,\boldsymbol{\beta}\right)\frac{1}{2}\frac{\partial^2 F}{\partial S_t^2} \\ &= (\theta_1+\theta_2 S_t)\lambda_i S_t^{\lambda_i-1}+0+\frac{1}{2}(\theta_3+\theta_4 S_t)\lambda_i\left(\lambda_i-1\right)S_t^{\lambda_i-2} \end{aligned}$$

and

$$L_1F\left(t_{k-1},S_{t_{k-1}},\lambda_i\right)=\sigma\left(t,S_t,\theta\right)\frac{\partial F}{\partial S_t}=\sqrt{\theta_3+\theta_4 S_t}\lambda_i S_t^{\lambda_i-1}\ .$$

$H_i\left(\boldsymbol{\beta},t,\Delta\right)$ can now be written as

$$\begin{aligned} H_i\left(\boldsymbol{\beta},t,\Delta\right) &= \frac{1}{n}\sum_{k=1}^{n}F_{k,i}^{(2)}\left(\boldsymbol{\beta}\right) \\ &= \frac{1}{n}\sum_{k=1}^{n}(Q_{\lambda_i,k,\Delta}-\left(L_1F\left(t_{k-1},S_{t_{k-1}},\lambda_i\right)\right)^2) \\ &= \frac{1}{n}\sum_{k=1}^{n}\left(Q_{\lambda_i,k,\Delta}-(\theta_3+\theta_4 S_{t_{k-1}})\lambda_i^2 S_{t_{k-1}}^{2\lambda_i-2}\right) \\ &= \frac{1}{n}\sum_{k=1}^{n}Q_{\lambda_i,k,\Delta}-\theta_3\lambda_i^2\frac{1}{n}\sum_{k=1}^{n}S_{t_{k-1}}^{2\lambda_i-2}-\theta_4\lambda_i^2\frac{1}{n}\sum_{k=1}^{n}S_{t_{k-1}}^{2\lambda_i-1} \end{aligned}$$

and

$$\begin{aligned}
G_i(\boldsymbol{\alpha},\boldsymbol{\beta},t,\Delta) &= \frac{1}{n}\sum_{k=1}^{n} F_{k,i}^{(1)}(\boldsymbol{\alpha},\boldsymbol{\beta}) \\
&= \frac{1}{n}\sum_{k=1}^{n}\left(D_{\lambda_i,k,\Delta} - L_0F\left(t_{k-1},S_{t_{k-1}},\lambda_i\right)\right) \\
&= \frac{1}{n}\sum_{k=1}^{n}(D_{\lambda_i,k,\Delta} \\
&\quad -(\theta_1+\theta_2 S_{t_{k-1}})\lambda_i S_{t_{k-1}}^{\lambda_i-1} + \frac{1}{2}(\theta_3+\theta_4 S_{t_{k-1}})\lambda_i\left(\lambda_i-1\right)S_{t_{k-1}}^{\lambda_i-2})) \\
&= \frac{1}{n}\sum_{k=1}^{n} D_{\lambda_i,k,\Delta} - \theta_1\lambda_i\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}^{\lambda_i-1} - \theta_2\lambda_i\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}^{\lambda_i} \\
&\quad -\frac{1}{2}\theta_3\lambda_i\left(\lambda_i-1\right)\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}^{\lambda_i-2} - \frac{1}{2}\theta_4\lambda_i\left(\lambda_i-1\right)\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}^{\lambda_i-1}.
\end{aligned}$$

Choose $\lambda_1=\lambda_3=1, \lambda_2=\lambda_4=2$. This leads to the following four equations:

$$\frac{1}{n}\sum_{k=1}^{n} D_{1,k,\Delta} - \theta_1 - \theta_2\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}} = 0\,,$$

$$\frac{1}{n}\sum_{k=1}^{n} D_{2,k,\Delta} - 2\theta_1\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}} - 2\theta_2\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}^2 - \theta_3 - \theta_4\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}} = 0\,,$$

$$\frac{1}{n}\sum_{k=1}^{n} Q_{1,k,\Delta} - \theta_3 - \theta_4\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}} = 0$$

and

$$\frac{1}{n}\sum_{k=1}^{n} Q_{2,k,\Delta} - 4\theta_3\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}^2 - 4\theta_4\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}^3 = 0.$$

This system has the solution

$$\widehat{\theta}_1 = \frac{1}{n}\sum_{k=1}^{n} D_{1,k,\Delta} - \theta_2\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}\,,$$

$$\widehat{\theta}_2 = \frac{\frac{1}{2}\left(\frac{1}{n}\sum_{k=1}^{n} D_{2,k,\Delta}\right) - \widehat{\theta}_3 - \widehat{\theta}_4\left(\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}\right) - \left(\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}\right)\left(\frac{1}{n}\sum_{k=1}^{n} D_{1,k,\Delta}\right)}{\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}^2 - \left(\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}\right)^2}\,,$$

$$\widehat{\theta}_3 = \frac{1}{n}\sum_{k=1}^{n} Q_{1,k,\Delta} - \widehat{\theta}_4 \frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}},$$

and

$$\widehat{\theta}_4 = \frac{\frac{1}{4}\left(\frac{1}{n}\sum_{k=1}^{n} Q_{2,k,\Delta}\right) - \left(\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}^2\right)\left(\frac{1}{n}\sum_{k=1}^{n} Q_{1,k,\Delta}\right)}{\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}^3 - \left(\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}\right)\left(\frac{1}{n}\sum_{k=1}^{n} S_{t_{k-1}}^2\right)}.$$

Hence

$$\widehat{\mu(t, S_t)} = \left(\widehat{\theta}_1 + \widehat{\theta}_2 S_t\right)$$

and

$$\widehat{\sigma(t, S_t)} = \sqrt{\widehat{\theta}_3 + \widehat{\theta}_4 S_t}.$$

Remark 3.4.1

In this chapter a parametric technique for measuring instantaneous volatility (for definition see Section 3.1) was given, based on solving a set of linear equations. These equations, with the only unknowns the parameters, were obtained via transform functions.

4

Realised Volatility

4.1 Introduction

As from this chapter the study will again focus mainly on volatility over larger time horizons, i.e daily volatility, weekly volatility etc. The main object is the prediction of the ex-ante expected volatility, but the techniques followed during the second stage were not satisfactory. The problem with volatility is that it is inherently unobservable, so, unlike any observable quantity, the stylised facts need to be found by somehow measuring volatility. The traditional measurements were based on models or were too noisy. **Researchers came to the realisation that before a forecasting model could be built, a volatility measurement (an ex-post measurement) that is non-parametric and error-free needs to be obtained**. Only then could all the characteristics of the volatility process be assessed and only then could an ex-ante expected volatility model be constructed that takes all the stylised facts into consideration.

A breakthrough came in volatility estimation, published by Anderson, Bollershev, Diebold and Labys (2001a,b), Barndoff-Nielsen and Shepard (2001), (2002a,b,c) and Comte and Renault (1998). They called this measurement realised volatility. This initiated the third stage. The essential difference between this approach and the approaches followed in the second stage is that instead of focusing on ex-ante expected volatility, researchers started focusing on ex-post volatility in a non-parametric way. Later in Corsi and Cursi (2003) another volatility measurement, called the Discrete Sine Transform (DST) approach, was given that works effectively. In Chapters 4 and 5, respectively, the realised volatility and the DST measurements are discussed in detail.

The realised volatility measurement is constructed using high-frequency return data. The main advantage, as is illustrated in this chapter, is that this measurement is model-free, and free of measurement error. The true volatility process need not be known when estimating the unobservable volatility. Volatility can now be seen as "observable". The characteristics of the distribution of volatility can readily be found and more reliable forecasting models can be created.

Some general definitions need to be given: As before let S_t be the market value of the share or index at time t with $p_t = \log(S_t)$. Assume that the share or index of interest is very marketable, so that at any instant a trade should occur. We can therefore model the return process and the volatility process in continuous time. The continuous return process is defined as $r_t = p_t - p_0$, with $t \in [0, T]$, i.e $S_t = S_0 e^{r_t}$.

A general return model is:

$$r_t = A_t + M_t, \tag{4.1.1}$$

where A_t denotes the predictable process, and M_t the unpredictable process (martingale).

The ex-post volatility measurement is derived from the above model. Some definitions and theorems are given in Section 4.2. In Section 4.3 the derivations of realised volatility are furnished. The unpredictable process can be decomposed into a continuous and a discrete part which is seen in real-life financial data.

4.2 Definitions and assumptions for derivations of realised volatility

In this section definitions and theorems are given that are to be used in Section 4.3 to obtain a non-parametric ex-post volatility measurement for r_t. The following definition comes from Anderson, Bollershev, Diebold and Labys (2001a).

Definition 4.2.1

For M_t, a martingale, the quadratic variation process $[M,M]_t$ is given by

$$[M,M]_t = M_t^2 - 2\int_0^t M_{s-}dM_s \, .$$

From the definition of an Ito integral it follows that

$$\lim_{n\to\infty} E\left[\sum_{k=1}^{n} M_{t_{k-1}}\left(M_{t_k} - M_{t_{k-1}}\right) - \int_0^t M_{s-}dM_s\right]^2 = 0 \, .$$

Mean square convergence implies convergence in probability
(denoted by $p\lim_{n\to\infty}$), i.e.

$$p\lim_{n\to\infty}\sum_{k=1}^{n} M_{t_{k-1}}\left(M_{t_k} - M_{t_{k-1}}\right) = \int_0^t M_{s-}dM_s \, .$$

The quadratic variation process is usually used as an ex-post measurement. In the following theorem the relationship between $[M,M]_t$ and $\sum_{k=1}^{n}\left(M_{t_k} - M_{t_{k-1}}\right)^2$ is derived when $n \to \infty$. This result is a very important building block in the build up to the realised volatility measurement.

Theorem 4.2.1

$$p\lim_{n\to\infty}\{\sum_{k=1}^{n}\left(M_{t_k} - M_{t_{k-1}}\right)^2\} = [M,M]_t \ where \, t_o = 0, t_n = t \, .$$

Proof

$$\begin{aligned}
p\lim_{n\to\infty}\{\sum_{k=1}^{n}\left(M_{t_k}-M_{t_{k-1}}\right)^2\} &= p\lim_{n\to\infty}\{\sum_{k=1}^{n}\left(M_{t_k}^2+M_{t_{k-1}}^2-2M_{t_k}M_{t_{k-1}}\right)\} \\
&= p\lim_{n\to\infty}\{\sum_{k=1}^{n}(M_{t_k}^2+M_{t_{k-1}}^2-2M_{t_{k-1}}(M_{t_k}-M_{t_{k-1}})- \\
&\quad 2M_{t_{k-1}}^2)\} \\
&= p\lim_{n\to\infty}\{\sum_{k=1}^{n}(M_{t_k}^2-M_{t_{k-1}}^2-2M_{t_{k-1}}(M_{t_k}-M_{t_{k-1}}))\} \\
&= M_t^2-2p\lim_{n\to\infty}\sum_{k=1}^{n}M_{t_{k-1}}(M_{t_k}-M_{t_{k-1}}) \text{ because } M_{t_o}=0 \\
&= M_t^2-2\int_0^t M_{s-}dM_s
\end{aligned}$$

which proves the theorem.

A practical implication of this result is that one can approximate the quadratic variation of r_t by taking the sum of the squared high-frequency returns. This is true because we can write

$$\begin{aligned}
[r,r]_t &= p\lim_{n\to\infty}\{\sum_{k=1}^{n}[r_{t_k}-r_{t_{k-1}}]^2\} \\
&= p\lim_{n\to\infty}\{\sum_{k=1}^{n}[p_{t_k}-p_o-(p_{t_{k-1}}-p_{o)}]^2\} \\
&= p\lim_{n\to\infty}\{\sum_{k=1}^{n}[p_{t_k}-p_{t_{k-1}}]^2\} \\
&= p\lim_{n\to\infty}\{\sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2\},
\end{aligned}$$

with $r_{t_k,t_{k-1}} = \log\frac{S_{t_k}}{S_{t_{k-1}}}$. The quadratic variation process $[r,r]_t$ can thus be approximated by $\sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2$, where $t_0 = 0$, $t_n = t$, and n is as large as possible. In the following theorem the equivalence of $[r,r]_t$ and $[M,M]_t$ is proved if model (4.1.1) is assumed.

Theorem 4.2.2

$$[r,r]_t = [M,M]_t \ ,$$

with

$$r_t = A_t + M_t \ ,$$

where A_t denotes the predictable process, and M_t the unpredictable process (martingale).

Proof

$$\begin{aligned}
[r,r]_t &= r_t^2 - 2\int_0^t r_{s-} dr_s \\
&= (M_t + A_t)^2 - 2\int_0^t (M_{s-} + A_{s-})\, dr_s \\
&= M_t^2 + A_t^2 + 2M_tA_t - 2\int_0^t M_{s-} dM_s - 2\int_0^t A_{s-} dM_s \\
&\quad -2\int_0^t M_{s-} dA_s - 2\int_0^t A_{s-} dA_s \\
&= M_t^2 - 2\int_0^t M_{s-} dM_s + 2\left(M_tA_t - \int_0^t A_{s-} dM_s - \int_0^t M_{s-} dA_s\right) \\
&\quad +A_t^2 - 2\int_0^t A_{s-} dA_s \\
&= [M,M]_t + 2\,[M,A]_t + [A,A]_t \ ,
\end{aligned}$$

with $[M,A]_t = p\lim_{n\to\infty}\{\sum_{k=1}^n [M_{t_k} - M_{t_{k-1}}][A_{t_k} - A_{t_{k-1}}]\}$, and $[A,A]_t = p\lim_{n\to\infty}\{\sum_{k=1}^n [A_{t_k} - A_{t_{k-1}}]^2\}$, with A_t predictable and of finite variation. Assuming that A_t is continuous, we have $A_{t_k} - A_{t_{k-1}} \approx 0 \Rightarrow [A,A]_t = 0$. By Cauchy-Schwartz it follows that

$$\begin{aligned}
& p\lim_{n\to\infty} \left| \{\sum_{k=1}^n [M_{t_k} - M_{t_{k-1}}][A_{t_k} - A_{t_{k-1}}]\} \right| \\
\leq \ & p\lim_{n\to\infty} \left[\sum_{k=1}^n \left(M_{t_k} - M_{t_{k-1}}\right)^2 \sum_{k=1}^n \left(A_{t_k} - A_{t_{k-1}}\right)^2 \right]^{\frac{1}{2}} \\
= \ & \left([M,M]_t\,[A,A]_t\right)^{\frac{1}{2}} \\
= \ & 0 \text{ since } [A,A]_t = 0.
\end{aligned}$$

Therefore:

$$[r, r]_t = [M, M]_t \ , \tag{4.2.1}$$

which proves the theorem.

4.3 Derivation of realised volatility

Definition 4.3.1

Realised volatility is defined as:

$$RV_{t,t-h} = \sum_{k=1}^{n} r_{t_k,t_{k-1}}^2, \tag{4.3.1}$$

where $t_0 = t - h$, $t_n = t$ and n is as large as possible.

Since

$$[r, r]_t - [r, r]_{t-h} = p \lim_{n\to\infty} \sum_{k=1}^{n} r_{t_k,t_{k-1}}^2,$$

$[r, r]_t - [r, r]_{t-h}$ may be approximated by the realised volatility measurement. In the following theorem an expression that suggests an estimator for the variation of the return process between times $t - h$ and t is derived. As such it is a very important result.

Theorem 4.3.1

$$Var\left(r_t | F_{t-h}\right) \approx E\left([r, r]_t - [r, r]_{t-h} | F_{t-h}\right)$$

over relatively small intervals. (See Section 3.2 for explanation of F_{t-h}.)

Proof

First consider:

$$\begin{aligned}
Var\left(r_t|F_{t-h}\right) &= E\left(\{r_t - E(r_t|F_{t-h})\}^2|F_{t-h}\right) \\
&= E(\{A_t + M_t - E(A_t + M_t|F_{t-h})\}^2|F_{t-h}) \\
&= E(\{A_t + M_t - E(A_t|F_{t-h}) - M_{t-h}\}^2|F_{t-h}) \\
&= E(\{A_t^2 + M_t^2 + (E(A_t|F_{t-h}))^2 + M_{t-h}^2 + \\
&\quad 2A_tM_t - 2A_tE(A_t|F_{t-h}) - 2A_tM_{t-h} - 2M_tE(A_t|F_{t-h}) \\
&\quad -2M_tM_{t-h} + 2E(A_t|F_{t-h})M_{t-h}\}|F_{t-h}) \\
&= E(A_t^2|F_{t-h}) + E(M_t^2|F_{t-h}) + (E(A_t|F_{t-h}))^2 + M_{t-h}^2 + \\
&\quad 2E(A_tM_t|F_{t-h}) - 2(E(A_t|F_{t-h}))^2 - 2E(A_tM_{t-h}|F_{t-h}) - \\
&\quad 2E(M_t|F_{t-h})E(A_t|F_{t-h}) - 2E(M_tM_{t-h}|F_{t-h}) + 2E(A_t|F_{t-h})M_{t-h} \\
&= E(M_t^2|F_{t-h}) - M_{t-h}^2 + (E(A_t^2|F_{t-h}) - (E(A_t|F_{t-h}))^2) + \\
&\quad 2(E(A_tM_t|F_{t-h}) - E(A_t|F_{t-h})M_{t-h}) \\
&= E(M_t^2|F_{t-h}) - M_{t-h}^2 + Var\left(A_t|F_{t-h}\right) + 2Cov(A_tM_t|F_{t-h})\,. \qquad (4.3.2)
\end{aligned}$$

The last two terms in (4.3.1) are negligible and have an influence only over long time horizons. Using Result (4.2.1) it follows that we may now also relate $E\left([r,r]_t - [r,r]_{t-h}|F_{t-h}\right)$ to the righthand side of (4.3.2):

$$\begin{aligned}
E\left([r,r]_t - [r,r]_{t-h}|F_{t-h}\right) &= E\left([M,M]_t - [M,M]_{t-h}|F_{t-h}\right) \\
&= E(M_t^2 - 2\int_0^t M_{s-}dM_s \\
&\quad -(M_{t-h}^2 - 2\int_0^{t-h} M_{s-}dM_s)|F_{t-h})) \\
&= E(M_t^2|F_{t-h}) - M_{t-h}^2 - 2E(\int_{t-h}^t M_{s-}dM_s|F_{t-h}) \\
&= E(M_t^2 \mid F_{t-h}) - M_{t-h}^2 \\
&\quad -2E(\int_{t-h}^t (M_{s-} - M_{t-h} + M_{t-h})dM_s|F_{t-h})
\end{aligned}$$

$$
\begin{aligned}
&= E(M_t^2|F_{t-h}) - M_{t-h}^2 - 2E(\int_{t-h}^{t}(M_{s-} - M_{t-h})dM_s|F_{t-h}) \\
&\quad -2E(\int_{t-h}^{t} M_{t-h}dM_s|F_{t-h}) \\
&= E(M_t^2|F_{t-h}) - M_{t-h}^2 - 2E(\int_{t-h}^{t}(M_{s-} - M_{t-h})dM_s) \\
&\quad -2E(M_{t-h}\int_{t-h}^{t} dM_s|F_{t-h}) \\
&= E(M_t^2 \mid F_{t-h}) - M_{t-h}^2 - 0 \\
&\quad -2E(M_{t-h}\left[M_t - M_{t-h}\right]|F_{t-h}) \\
&= E(M_t^2 \mid F_{t-h}) - M_{t-h}^2 - 0 \\
&\quad -2E(M_{t-h}M_t - M_{t-h}^2|F_{t-h}) \\
&= E(M_t^2 \mid F_{t-h}) - M_{t-h}^2 - 0 \\
&\quad -2M_{t-h}E(M_t|F_{t-h}) + 2M_{t-h}^2 \\
&= E(M_t^2|F_{t-h}) - M_{t-h}^2 \,.
\end{aligned}
$$

We have that

$$Var\,(A_t|F_{t-h}) \approx Cov(A_tM_t|F_{t-h}) \approx 0\,,$$

so

$$
\begin{aligned}
E\left([r,r]_t - [r,r]_{t-h}\,|F_{t-h}\right) &= E(M_t^2|F_{t-h}) - M_{t-h}^2\,. \\
&= Var\,(r_t|F_{t-h})\;. \qquad (4.3.3)
\end{aligned}
$$

This proves the theorem.

Remark 4.3.1

$$[r,r]_t - [r,r]_{t-h} \qquad (4.3.4)$$

is approximately an unbiased estimator for

$$Var\,(r_t|F_{t-h})\,,$$

for small h.

In practice $Var\,(A_t|F_{t-h}) \approx Cov(A_t M_t|F_{t-h}) \approx 0$ and differs from 0 only over long-time horizons. The reason why (4.3.3) is so important is that while Var$(r_t \mid F_{t-h})$ is unobservable, (4.3.4) is actually observable in the sense that we can approximate (4.3.4) by

$$\sum_{k=1}^{n} [r_{t_k,t_{k-1}}]^2,$$

where $t_o = t - h$, and $t_n = t$.

After obtaining estimates for Var$(r_t|F_{t-h})$, the stylised facts of

$$[r, r]_t - [r, r]_{t-h}$$

can be studied. A model for Var$(r_t \mid F_{t-h})$ for forecasting purposes can then be built. Remember that

$$\sum_{k=1}^{n} [r_{t_k,t_{k-1}}]^2,$$

where $t_0 = t - h$, and $t_n = t$ converges to (4.3.4) and not to

$$E\left([r, r]_t - [r, r]_{t-h} |F_{t-h}\right).$$

We actually want

$$Var\,(r_t|F_{t-h}) - E\left(\sum_{k=1}^{n} [r_{t_k,t_{k-1}}]^2 |F_{t-h}\right) = o(1),$$

but we have only shown that

$$Var\,(r_t \mid F_{t-h}) = E\left([r, r]_t - [r, r]_{t-h} |F_{t-h}\right)$$

and

$$[r, r]_t - [r, r]_{t-h} = p \lim_{n\to\infty} \sum_{k=1}^{n} [r_{t_k,t_{k-1}}]^2,$$

where $t_0 = t - h$ and $t_n = t$.

The problem is however that convergence in probability does not imply convergence in expectation. Under some additional weak assumptions, we will have convergence in expectation too. These assumptions are addressed in Theorem 4.3.2.

Theorem 4.3.2

If

$$r_t = M_t + A_t,$$

with M_t a martingale, and A_t satisfies the following conditions:

i) A_t is continuous,

ii) Var$(A_t \mid F_{t-h})$ =Cov$(A_t M_t \mid F_{t-h}) = 0$,

iii) $\mathrm{E}\left(\sum_{k=1}^{n}\left(A_{t_k} - A_{t_{k-1}}\right)^2 \mid F_{t-h}\right) \longrightarrow 0$, and

iv) $\mathrm{E}\left(\sum_{k=1}^{n}\left(A_{t_k} - A_{t_{k-1}}\right)\left(M_{tk} - M_{t_{k-1}}\right) \mid F_{t-h}\right) \longrightarrow 0$ as $n \to \infty$.

Then we have

$$Var\left(r_t | F_{t-h}\right) - E\left(\sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2 | F_{t-h}\right) = o(1)\ ,$$

with $t_o = t - h$, and $t_n = t$.

Proof

From (4.2.1) it follows that

$$E([r,r]_t - [r,r]_{t-h}\,|F_{t-h}) = E([M,M]_t - [M,M]_{t-h}\,|F_{t-h})\ . \tag{4.3.5}$$

Further for $t_o = t - h$, and $t_n = t$ it follows that

$$\begin{aligned}\sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2 &= \sum_{k=1}^{n}[A_{t_k} + M_{t_k} - \left(A_{t_{k-1}} + M_{t_{k-1}}\right)]^2 \\ &= \sum_{k=1}^{n}\left(A_{t_k} - A_{t_{k-1}}\right)^2 + 2\sum_{k=1}^{n}\left(A_{t_k} - A_{t_{k-1}}\right)\left(M_{tk} - M_{t_{k-1}}\right) + \\ &\quad \sum_{k=1}^{n}\left(M_{t_k} - M_{t_{k-1}}\right)^2\end{aligned}$$

$$\begin{aligned}
\Rightarrow \quad & E\left(\sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2|F_{t-h}\right) = E\left(\sum_{k=1}^{n}\left(A_{t_k}-A_{t_{k-1}}\right)^2|F_{t-h}\right) + \\
& 2E\left(\sum_{k=1}^{n}\left(A_{t_k}-A_{t_{k-1}}\right)\left(M_{t_k}-M_{t_{k-1}}\right)|F_{t-h}\right) + \\
& E\left(\sum_{k=1}^{n}\left(M_{t_k}-M_{t_{k-1}}\right)^2|F_{t-h}\right) && (4.1) \\
\rightarrow \quad & E\left(\sum_{k=1}^{n}\left(M_{t_k}-M_{t_{k-1}}\right)^2|F_{t-h}\right) \text{ as n} \rightarrow \infty, && (4.3.6) \\
\text{with } t_o \quad = \quad & t-h \text{ and } t_n = t\,.
\end{aligned}$$

From (4.3.3) it follows that

$$\begin{aligned}
E\left([r,r]_t - [r,r]_{t-h}|F_{t-h}\right) &= E\left([M,M]_t - [M,M]_{t-h}|F_{t-h}\right) \\
&= E(M_t^2|F_{t-h}) - M_{t-h}^2 \\
&= E(M_t^2 - M_{t-h}^2|F_{t-h}) \\
&= E\left(\left(\sum_{k=1}^{n}(M_{t_k}-M_{t_{k-1}})\right)^2|F_{t-h}\right) \\
&\quad -E\left(\sum_{k=1}^{m}(M_{t_k}-M_{t_{k-1}}))^2|F_{t-h}\right),
\end{aligned}$$

where $t_o = 0, t_n = t$ and $t_m = t-h$.

$$\begin{aligned}
E\left([r,r]_t - [r,r]_{t-h}|F_{t-h}\right) &= E(\sum_{k=1}^{n}M_{t_k}^2 - 2\sum_{k=1}^{n}M_{t_k}M_{t_{k-1}} + \sum_{k=1}^{n}M_{t_{k-1}}^2 \\
&\quad -\sum_{k=1}^{m}M_{t_k}^2 + 2\sum_{k=1}^{m}M_{t_k}M_{t_{k-1}} - \sum_{k=1}^{m}M_{t_{k-1}}^2|F_{t-h}) \\
&= E\left(\sum_{k=m+1}^{n}M_{t_k}^2 - 2\sum_{k=m+1}^{n}M_{t_k}M_{t_{k-1}} + \sum_{k=m+1}^{n}M_{t_{k-1}}^2|F_{t-h}\right) \\
&= E\left(\sum_{k=m+1}^{n}(M_{t_k}-M_{t_{k-1}})^2|F_{t-h}\right) \\
&= E\left(\sum_{k=1}^{n}\left(M_{t_k}-M_{t_{k-1}}\right)^2|F_{t-h}\right), && (4.3.7)
\end{aligned}$$

with $t_o = t - h, t_n = t$.

Thus from (4.3.6) and (4.3.7) we have that

$$\begin{aligned} Var\left(r_t \mid F_{t-h}\right) &= E([r,r]_t - [r,r]_{t-h} \mid F_{t-h}) \\ &= \lim_{n\to\infty} E\left(\sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2 \mid F_{t-h}\right), \end{aligned}$$

which proves the theorem.

The above assumptions are weak and in practice hold for most cases, at least over short-time horizons. The main result in this section is that if the return process is continuous and r_t can be modelled as

$$r_t = M_t + A_t,$$

then the ex-post (actual) variance of the return process between time $t-h$ and t can be approximated by

$$\sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2,$$

where $t_o = t - h$, and $t_n = t$.

4.4 Decomposing the unpredictable part

In practice the return process usually consists of jumps caused by rare events. Till now, no assumptions were made as to whether the unpredictable part is continuous or discrete. Consider $r_t = M_t + A_t$, where the martingale part of r_t is decomposed into a continuous part and a jump component part. This decomposition is seen in real-world returns. It follows that

$$r_t = M_t^C + M_t^J + A_t\,,$$

where M_t^C denotes the continuous part, and M_t^J denotes the jump component part. We have

$$M_t = M_t^C + M_t^J.$$

In Section 4.3 the characteristics of $M_t^C + M_t^J$ $(= M_t)$ were examined. In this section the characteristics of the separate parts will be dealt with.

Let M_t^J be the sum of all the jumps until time t. Let $jump_{t_i}$ denote the the jump size at time $t_i \in [0,t]$, and N be the number of jumps in time interval $[0,t]$. Here the jump size, jump times and number of jumps may be random variables. So,

$$M_t^J = \sum_{i=1}^{N} jump_{t_i}.$$

We can now derive the quadratic variation process of r_t and see what it looks like in terms of M_t^C and M_t^J. We will then show how the separate parts can be approximated.

Theorem 4.4.1

$$[r,r]_t = \left[M^C, M^C\right]_t + \sum_{i=1}^{N} jump_{t_i}^2 .$$

Proof

$$\begin{aligned}
[r,r]_t &= [M,M]_t \\
&= \left[M^C + M^J, M^C + M^J\right]_t \\
&= \left(M_t^C + M_t^J\right)^2 - 2\int_0^t (M_{s-}^C + M_{s-}^J) dM_s \\
&= (M_t^C)^2 + (M_t^J)^2 + 2M_t^C M_t^J - 2\int_0^t M_{s-}^C dM_s^C \\
&\quad -2\int_0^t M_{s-}^C dM_s^J - 2\int_0^t M_{s-}^J dM_s^C - 2\int_0^t M_{s-}^J dM_s^J,
\end{aligned}$$

Since M^J is discrete, the integrals can be replaced by summations.
Thus

$$\begin{aligned}
[r,r]_t &= (M_t^C)^2 - 2\int_0^t M_{s-}^C dM_s^C + (\sum_{i=1}^{N} jump_{t_i})^2 + \\
&\quad 2M_t^C M_t^J - 2\sum_{i=1}^{N} M_{t_i}^C jump_{t_i} - 2\sum_{i=1}^{N} M_{t_i}^J (M_{t_{i+1}}^C - M_{t_i}^C) \\
&\quad -2\sum_{i=1}^{N} M_{t_{i-1}}^J jump_{t_i} \text{, where } M_{t_{N"+1}}^C \equiv M_t^C
\end{aligned} \tag{4.4.1}$$

because $M_{t_i}^J$ is constant between consecutive jumps and $dM_{t_i}^J = M_{t_i}^J - M_{t_{i-1}}^J$ is just $jump_{t_i}$.

Consider now the third, fourth and seventh terms in (4.4.1), i.e.

$$(\sum_{i=1}^{N} jump_{t_i})^2 = \sum_{i=1}^{N} jump_{t_i}^2 + 2\sum_{i<j}\sum jump_{t_i} jump_{t_j}$$

and

$$\begin{aligned}
M_t^C M_t^J &= \left(M_{t_1}^C + \{M_{t_2}^C - M_{t_1}^C\} + ... + \{M_{t_N}^C - M_{t_{N-1}}^C\} + \{M_t^C - M_{t_N}^C\}\right) \cdot \\
&\quad (jump_{t_1} + jump_{t_2} + ... + jump_{t_N}) \\
&= (M_{t_1}^C jump_{t_1} + M_{t_2}^C jump_{t_2} + ... + M_{t_N}^C jump_{t_N}) + \\
&\quad [(M_{t_2}^C - M_{t_1}^C) jump_{t_1} + (M_{t_3}^C - M_{t_2}^C)(jump_{t_1} + jump_{t_2}) + ... \\
&\quad +(M_{t_N}^C - M_{t_{N-1}}^C)(jump_{t_1} + jump_{t_2} + ... + jump_{t_{N-1}}) \\
&\quad + \left(M_t^C - M_{t_N}^C\right)(jump_{t_1} + jump_{t_2} + ... + jump_{t_N})] \\
&\quad + \sum_{i=2}^{N} jump_{t_i} \sum_{i>j} (M_{t_j}^C - M_{t_j}^C) \\
&= \sum_{i=1}^{N} M_{t_i}^C jump_{t_i} + [(M_{t_2}^C - M_{t_1}^C) M_{t_1}^J + (M_{t_3}^C - M_{t_2}^C) M_{t_2}^J + ... \\
&\quad +(M_{t_N}^C - M_{t_{N-1}}^C) M_{t_{N-1}}^J + \left(M_t^C - M_{t_N}^C\right) M_{t_N}^J] + 0 \\
&= \sum_{i=1}^{N} M_{t_i}^C jump_{t_i} + \sum_{i=1}^{N} M_{t_i}^J (M_{t_{i+1}}^C - M_{t_i}^C)
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i=1}^{N} M_{t_{i-1}}^J jump_{t_i} &= (jump_{t_1} jump_{t_2} + \{jump_{t_1} + jump_{t_2}\} jump_{t_3} + ... + \\
&\quad \{jump_{t_1} + jump_{t_2} + ... + jump_{t_{N-1}}\} jump_{t_N}) \\
&= (\{jump_{t_1} jump_{t_2} + jump_{t_1} jump_{t_3} + ... + jump_{t_1} jump_{t_N}\} + \\
&\quad \{jump_{t_2} jump_{t_3} + jump_{t_2} jump_{t_4} + ... \\
&\quad + jump_{t_2} jump_{t_N}\} + ... + \{jump_{t_{N-1}} jump_{t_N}\}) \\
&= \sum_{i<j}\sum jump_{t_i} jump_{t_j}.
\end{aligned}$$

Therefore

$$\begin{aligned}
[r,r]_t &= \left[M^C,M^C\right]_t + \sum_{i=1}^{N} jump_{t_i}^2 + 2\sum_{i<j}\sum jump_{t_i}jump_{t_j} + 2(\sum_{i=1}^{N} M_{t_i}^C jump_{t_i} \\
&\quad + \sum_{i=1}^{N} M_{t_i}^J(M_{t_{i+1}}^C - M_{t_i}^C)) - 2\sum_{i=1}^{N} M_{t_i}^C jump_{t_i} \\
&\quad -2\sum_{i=1}^{N} M_{t_i}^J(M_{t_{i+1}}^C - M_{t_i}^C) - 2\sum_{i<j}\sum jump_{t_i}jump_{t_j} \\
&= \left[M^C,M^C\right]_t + \sum_{i=1}^{N} jump_{t_i}^2 ,
\end{aligned}$$

which proves the theorem.

We can write $\sum_{i=1}^{N} jump_{t_i}$ in more detail as

$$\sum_{i=1}^{N} jump_{t_i} = \sum_{0\leq t_s \leq t} \kappa_{t_s}\Delta J_{t_s} , \tag{4.4.2}$$

where κ_s is a random variable denoting the size of the jump, and

$$\begin{aligned}
\Delta J_{t_s} &= J_{t_s} - J_{t_{s-1}} \\
J_{t_s} &= N_{t_s} - \lambda \cdot t_s ,
\end{aligned}$$

where N_{t_s} is a Poisson process, and λ the rate of the process. Therefore

$$\begin{aligned}
\Delta J_{t_s} &= N_{t_s} - \lambda\cdot t_s - N_{t_{s-1}} + \lambda \cdot t_{s-1} \\
&= \left(N_{t_s} - N_{t_{s-1}}\right) - \lambda\left(t_s - t_{s-1}\right) \\
&= \underset{in\,(t_s - t_{s-1})}{\#Jumps} - \underset{in\,(t_s-t_{s-1})}{E(\#Jumps)}
\end{aligned}$$

Barndorff-Nielsen and Shepard (2003) proved the following important result that we give here without proof.

$$\left[M^C,M^C\right]_t = \lim_{n\to\infty}\frac{\pi}{2}\sum_{k=1}^{n} \mid r_{t_k,t_{k-1}} \mid\mid r_{t_{k-1},t_{k-2}} \mid \equiv RB_t, .$$

where $t_k \in [0,t]$ for all k and RB_t is the *Realised Bipower* variation measurement.

From Theorem 4.4.1 and (4.4.2) it follows that

$$[r, r]_t = \left[M^C, M^C\right]_t + \sum_{0 \leq t_s \leq t} \kappa_{t_s}^2 \Delta N_{t_s}^2 \,. \tag{4.4.3}$$

So $\sum_{0 \leq t_s \leq t} \kappa_{t_s}^2 \Delta N_{t_s}^2$ may be approximated by

$$RJ_t = \sum_{k=1}^{n} [r_{t_k, t_{k-1}}]^2 - \frac{\pi}{2} \sum_{k=1}^{n} | \, r_{t_k, t_{k-1}} \, || \, r_{t_{k-1}, t_{k-2}} \, | = RV_t - RB_t \text{ , with } t_k \in [0, t]$$

if this term is positive; otherwise approximate $\sum_{0 \leq t_s \leq t} \kappa_{t_s}^2 \Delta N_{t_s}^2$ by zero. RJ_t denotes the realised jump variation measurement.

4.5 Summary

In this chapter a non-parametric, ex-post volatility measurement was given, based on high-frequency return data, using quadratic variation theory. In theory, the more frequently the return data is sampled, the closer the realised volatility (defined as the sum of squared ex-post high-frequency returns) will become to the actual volatility. **In this sense volatility can be seen as observable**. Using this measurement, the stylised facts of volatility can then be identified. The most important results of Chapter 4 are:

i) Under the assumption that the share or index of interest is very marketable, so that at any instant a trade should occur, we may model the return process and the volatility process in continuous time. The continuous return process is therefore defined as $r_t = p_t - p_0$, with $t \in [0, T]$, i.e $S_t = S_0 e^{r_t}$.

   A general return model is:

$$r_t = A_t + M_t,$$

   where A_t denotes the predictable process, and M_t the unpredictable process (martingale).

ii) The quadratic variation process of r_t (Definition 4.2.1) can be approximated by taking the sum of squared ex-post high-frequency returns.

iii) The variance of r_t is equal to the expected value of the quadratic variation of r_t over the same time-horizon.

iv) Under the weak assumptions given in Theorem 4.2.2, the sum of squared ex-post returns over a given time-horizon converges to the variance of r_t over the same time-horizon.

5

The discrete sine transform measurement

5.1 Introduction

The realised volatility measurement uses high-frequency returns. The smaller the time interval of the returns, the better the measurement. Conversely, when too high-frequency data is used, performance of the volatility measurement is adversely influenced by microstructure effects (explained in Section 5.2). There is thus an optimal point where sampling more frequent returns actually makes the realised volatility measurement perform worse. In this chapter a volatility measurement (termed the DST measurement) which overcomes the shortcomings of the realised volatility is given and discussed. In Chapter 7 the shortcomings of both the realised volatility measurement and the DST measurement are discussed and a volatility measurement that overcomes the shortcomings of these measurements is given. Section 5.2 explains why the realised volatility measurement fails under certain conditions, Section 5.3 shows the derivation of the DST meaurement and Section 5.4 shows how to obtain the DST measurement in practice. This measurement was introduced in Corsi and Cursi (2003).

5.2 Practical problems when using the realised volatility method

Before some practical problems are considered, microstructure effects of high-frequency market returns are discussed briefly. Owing to microstructure effects, very high-frequency returns are negatively autocorrelated. The two main components of microstructure effects are the bid-ask bounce and price discreteness. These two components result in negatively autocorrelated high-frequency returns. To illustrate the effect of the bid-ask bounce, consider the following quotation from Roll (1984):

"*Let A≡ask price, B≡bid price and suppose that the successive transactions are equally likely to be a purchase or a sale. Suppose at time* $t-1$*, the price is a sale to the market maker and therefore the price is at B.*

Now at time t*, the possible prices can either be A or B with expected price*$=\frac{A+B}{2} > B$ *. If on the other hand the price at time* $t-1$ *is A, then at time* t *again we have that expected price*$=\frac{A+B}{2} < A$*.Thus resulting in a bid-ask bounce . That is, a higher price will be followed by an expected lower price, and a lower price will be followed by an expected higher price, inducing negative first order autocorrelation.*

Price discreteness also induce negative first order autocorrelation. The observable price is obtained by rounding the underlying true values. Thus if price is rounded down will tend to go up again and if rounded up will tend to go down again."

The problem with realised volatility is that if the index or stock is not highly marketable, the assumption that the price process is continuous will not hold. Furthermore, if the stock or index is not very marketable, microstructure effects may become problematic. This method assumes that returns are independent. If microstructure effects are evident, then high-frequency returns are not independent, with the result that the variance of the sum is not the sum of the variances. So there is a trade-off between sampling at high frequencies to get good approximations for the quadratic variation of returns and the microstructure effects. For daily volatility measurements estimates we need to sample at least every 5 minutes (288 values per day) to get a reasonable measurement error. This obviously depends on the time horizon over which the volatility estimates are needed. So, if microstructure effects are still evident in the 5 minute data and the time frame is relatively small, one needs to look at other measurements that take the microstructure effect into account. In very liquid markets, 5 minute returns have no microstructure effects. [See Anderson, Borrerslev, Diebold and Labys, 2001a.] However, in less liquid markets, such as the South African market, even daily returns may be correlated.

5.3 The discrete sine transform method

In Corsi and Cursi (2003) the following model was proposed to incorporate first order autocorrelation. As we will be illustrating that this measurement works better than the realised volatility measurement in the existence of microstructure effects, the tick times will not be modelled. The assumption is made that the time between trades is constant.

As usual the price of the stock or index is decomposed into a martingale component and an error term, expressing the difference between the observed price and the martingale component. The observed logarithmic price at tick time t_n is then given by:

$$\log(S_{t_n}) = p_{t_n} = \tilde{p}_{t_n} + \eta\omega_{t_n} \ ,$$

where $\tilde{p}_{t_n}$ is the martingale component, and $\eta\omega_{t_n}$ the error term. Hence:

$$r_{t_n} = p_{t_n} - p_{t_{n-1}} = \tilde{p}_{t_n} + \eta\omega_{t_n} - \tilde{p}_{t_{n-1}} - \eta\omega_{t_{n-1}} = (\tilde{p}_{t_n} - \tilde{p}_{t_{n-1}}) + \eta(\omega_{t_n} - \omega_{t_{n-1}}) \ ,$$

which can be written as:

$$r_{t_n} = \sigma\tilde{\epsilon}_{t_n} + \eta(\omega_{t_n} - \omega_{t_{n-1}}) \ ,$$

with $\tilde{\epsilon}_{t_n} \sim IID(0;1), \omega_{t_n} \sim IID(0,1)$ and $\tilde{\epsilon}_{t_n}$ and ω_{t_n} are uncorrelated.

This is an MA(1) model, with $\mathrm{E}(r_{t_n}) = 0$. Also,

$$\begin{aligned} E\left(r_{t_n}^2\right) &= E\left(\sigma\tilde{\epsilon}_{t_n} + \eta(\omega_{t_n} - \omega_{t_{n-1}})\right)^2 \\ &= E\left(\sigma^2\tilde{\epsilon}_{t_n}^2 + \eta^2(\omega_{t_n} - \omega_{t_{n-1}})^2 + 2\sigma\eta\tilde{\epsilon}_{t_n}(\omega_{t_n} - \omega_{t_{n-1}})\right) \\ &= \sigma^2 Var\left(\tilde{\epsilon}_{t_n}\right) + \eta^2 Var\left(\omega_{t_n} - \omega_{t_{n-1}}\right) \\ &= \sigma^2 + 2\eta^2 \end{aligned}$$

and

$$
\begin{aligned}
E\left(r_{t_n} r_{t_{n-1}}\right) &= E\left(r_{t_n} r_{t_{n+1}}\right) \\
&= E\left(\sigma\tilde{\epsilon}_{t_n} + \eta(\omega_{t_n} - \omega_{t_{n-1}})\right)\left(\sigma\tilde{\epsilon}_{t_{n+1}} + \eta(\omega_{t_{n+1}} - \omega_{t_n})\right) \\
&= E(\sigma^2\tilde{\epsilon}_{t_n}\tilde{\epsilon}_{t_{n+1}} + \sigma\tilde{\epsilon}_{t_n}\eta(\omega_{t_{n+1}} - \omega_{t_n}) + \sigma\tilde{\epsilon}_{t_{n+1}}\eta(\omega_{t_n} - \omega_{t_{n-1}}) + \\
&\quad \eta^2(\omega_{t_n} - \omega_{t_{n-1}})(\omega_{t_{n+1}} - \omega_{t_n})) \\
&= 0 + 0 + 0 + E\left(\eta^2(\omega_{t_n} - \omega_{t_{n-1}})(\omega_{t_{n+1}} - \omega_{t_n})\right) \\
&= -\eta^2 E\left(\omega_{t_n}^2\right) \\
&= -\eta^2 .
\end{aligned}
$$

MA(1) processes arise naturally in microstructure models of tick-by-tick returns and the non-parametric DST measurement decorrelates signal for data exhibiting MA(1) type of behaviour.

Here is how it works: Suppose we have a set of returns of size m.
Let $\mathbf{r} = \left[r_{t_n}, r_{t_{n-1}}, ..., r_{t_{n-m+1}}\right]^T$ and $\Sigma : m \times m$ the associated covariance matrix. Then

$$
\Sigma : m \times m = \begin{bmatrix}
\sigma^2 + 2\eta^2 & -\eta^2 & 0 & \cdots & 0 & 0 & 0 \\
-\eta^2 & \sigma^2 + 2\eta^2 & -\eta^2 & \cdots & 0 & 0 & 0 \\
 & \ddots & \ddots & \ddots & & & \\
 & & \ddots & \ddots & \ddots & & \\
 & & & \ddots & \ddots & \ddots & \\
0 & 0 & 0 & \cdots & -\eta^2 & \sigma^2 + 2\eta^2 & -\eta^2 \\
0 & 0 & 0 & \cdots & 0 & -\eta^2 & \sigma^2 + 2\eta^2
\end{bmatrix} .
$$

Let $(\lambda_i, \mathbf{e}_i)$ $i = 1, 2, ..., m$ be the eigenvalue-eigenvector pairs associated with Σ.

Theorem 5.2.1

$$
\lambda_i = \sigma^2 + 4\eta^2 \sin^2\left(\frac{\pi(m + 1 - i)}{2(m + 1)}\right) \text{ for } i = 1, 2, ..., m \tag{5.3.1}
$$

and

$$
\begin{aligned}
\mathbf{e}_i &= \{\sqrt{\frac{2}{m+1}} \sin\left(\frac{\pi\,(m+1-i)\,k}{m+1}\right)\} \text{ for } k = 1,2,...,m \\
&= \sqrt{\frac{2}{m+1}} \left[\sin(\frac{\pi\,(m+1-i)}{m+1}), \sin(\frac{2\pi\,(m+1-i)}{m+1}), ..., \sin(\frac{\pi m\,(m+1-i)}{m+1})\right]^T
\end{aligned}
$$

with $\lambda_1 > \lambda_2 > ... > \lambda_m$.

Proof

Looking at (5.3.1) we see that the λ_i 's for $i = 1, 2, ...m$ rotate around $\sin^2\left(\frac{\pi}{4}\right)$. It will be shown that this is indeed the case by explicitly deriving the eigenvalues for $m = 1$ to $m = 4$. The eigenvectors of Σ will not be derived.

Let $\Sigma^* = \Sigma - \sigma^2 I_m$. Then the eigenvalues of Σ^* follow from the solution of $\det(\Sigma^* - \lambda^* I_m) = 0$, i.e

$$
\begin{aligned}
\det(\Sigma^* - \lambda^* I_m) &= D_m \\
&= \det \begin{bmatrix}
2\eta^2-\lambda^* & -\eta^2 & 0 & \cdots & 0 & 0 & 0 \\
-\eta^2 & 2\eta^2-\lambda^* & -\eta^2 & \cdots & 0 & 0 & 0 \\
 & \ddots & \ddots & \ddots & & & \\
 & & \ddots & \ddots & \ddots & & \\
 & & & \ddots & \ddots & \ddots & \\
0 & 0 & 0 & \cdots & -\eta^2 & 2\eta^2-\lambda^* & -\eta^2 \\
0 & 0 & 0 & \cdots & 0 & -\eta^2 & 2\eta^2-\lambda^*
\end{bmatrix}. \\
&= (2\eta^2 - \lambda^*) D_{m-1} - \eta^4 D_{m-2}
\end{aligned}
$$

with m the dimension of the matrix.

Let $X = 2\eta^2 - \lambda^*$ and $c = \eta^2$ .Then it follows that

$$D_m = X D_{m-1} - c^2 D_{m-2}\,.$$

Hence for $m = 1$:

$$D_1 = 2\eta^2 - \lambda^* = X$$

and

$$2\eta^2 - \lambda^* = 0 \Rightarrow \lambda^* = 2\eta^2 = 4\eta^2 \left(\frac{1}{2}\right).$$

for $m = 2$:

$$D_2 = X^2 - c^2 = 0$$

$$\Rightarrow X = \pm c$$

and

$$\lambda^* = 2\eta^2 - X = 2\eta^2 \pm c = 2\eta^2 \pm \eta^2 = 4\eta^2(\frac{1}{2} \pm \frac{1}{2^2}).$$

for $m = 3$:

$$D_3 = X(X^2 - c^2) - c^2 X = X(X^2 - 2c^2) = 0$$

$$X = 0 \text{ or } \pm\sqrt{2}c$$

$$\begin{aligned}
&\Rightarrow \quad \lambda^* = 2\eta^2 \text{ or } 2\eta^2 \pm \sqrt{2}c \\
&\Rightarrow \quad \lambda^* = 2\eta^2 \text{ or } 2\eta^2 \pm \sqrt{2}\eta^2 \\
&\Rightarrow \quad \lambda^* = 2\eta^2 \pm \sqrt{2}\eta^2 \\
&\Rightarrow \quad \lambda^* = 2\eta^2 \text{ or } 4\eta^2(\frac{1}{2} \pm \frac{1}{2^{\frac{3}{2}}}).
\end{aligned}$$

for $m = 4$:

$$D_4 = X(X^3 - 2c^2 X) - c^2(X^2 - c^2) = X^4 - 3c^2 X^2 + c^4 = 0$$

$$\begin{aligned}
\Rightarrow \quad X^2 &= \frac{3c^2 \pm \sqrt{9c^4 - 4c^4}}{2} \\
&= \frac{3}{2}c^2 \pm \frac{\sqrt{5}}{2}c^2 \\
&= \left(\frac{3}{2} \pm \frac{\sqrt{5}}{2}\right)c^2 \\
&= \frac{3 \pm \sqrt{5}}{2}c^2
\end{aligned}$$

$$\Rightarrow X = \pm\sqrt{\frac{3 \pm \sqrt{5}}{2}}c$$

$$\begin{aligned}
\Rightarrow \quad \lambda^* &= 2\eta^2 \pm \sqrt{\frac{3 \pm \sqrt{5}}{2}}\eta^2 \\
&= \eta^2 \left(2 \pm \sqrt{\frac{3 \pm \sqrt{5}}{2}}\right) \\
&= 4\eta^2(\frac{1}{2} \pm \frac{\sqrt{3+\sqrt{5}}}{2^{\frac{5}{2}}}) \text{ or } 4\eta^2(\frac{1}{2} \pm \frac{\sqrt{3-\sqrt{5}}}{2^{\frac{5}{2}}})\ .
\end{aligned}$$

We see that

$$\begin{aligned}
\lambda_i^* &= 4\eta^2(\frac{1}{2} \pm constant) \\
&= 4\eta^2(\sin^2\left(\frac{\pi}{4}\right) \pm constant)\ ,
\end{aligned}$$

with $i = 1, 2, ..., m$.

This completes the proof.

According to principal component analysis: $Y_i = \mathbf{e}_i^T \mathbf{r}$ is the i-th principal component, and

$$\begin{aligned}
Var(Y_i) &= \lambda_i = \sigma^2 + \lambda_i^* \\
&= \sigma^2 + 4\eta^2 \sin^2\left(\frac{\pi\,(m+1-i)}{2\,(m+1)}\right) \text{ for } i = 1, 2, ..., m.
\end{aligned}$$

We are interested in the principal component Y_m associated with the smallest eigenvalue because it has the smallest variance, i.e. the best estimator

$$Var(Y_m) = \lambda_m = \sigma^2 + 4\eta^2 \sin^2\left(\frac{\pi}{2\,(m+1)}\right)$$

and

$$\lim_{m\to\infty} Var(Y_m) = \lim_{m\to\infty} \lambda_m = \sigma^2.$$

The smallest eigenvalue of Σ if $m \to \infty$ is therefore equal to the volatility σ^2 between consecutive tick times. For large m the smallest eigenvalue of Σ, i.e. λ_m, can approximately be decomposed into the signal component (the martingale component) and error component

$$\lambda_m \simeq \sigma^2 + 4\eta^2 \frac{\pi^2}{4(m+1)^2} = \sigma^2 + \frac{\eta^2\pi^2}{(m+1)^2}$$
$$\text{because } \left(\sin^2(x) \simeq x^2 \text{ for small } x\right).$$

The first term σ^2 is due to the martingale component and the second term $\frac{\eta^2\pi^2}{(m+1)^2}$ to the error term.

5.4 How to obtain the DST measurement in practice

The DST measurement is relatively easy to obtain. Suppose there are $m \times n$ tick times, i.e. $m \times n$ returns. To obtain an estimate of the volatility σ^2 between ticktimes, place every m returns into a vector $\mathbf{r}_i = \left[r_{t_i}, r_{t_{i-1}}, ..., r_{t_{i-m+1}}\right]^T$ for $i = 1, 2, ..., n$, with $\Sigma : m \times m$ the associated covariance matrix.

$$\Sigma : m \times m = \begin{bmatrix} \sigma^2 + 2\eta^2 & -\eta^2 & 0 & \cdots & 0 & 0 & 0 \\ -\eta^2 & \sigma^2 + 2\eta^2 & -\eta^2 & \cdots & 0 & 0 & 0 \\ & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ 0 & 0 & 0 & \cdots & -\eta^2 & \sigma^2 + 2\eta^2 & -\eta^2 \\ 0 & 0 & 0 & \cdots & 0 & -\eta^2 & \sigma^2 + 2\eta^2 \end{bmatrix}$$

and

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} (\mathbf{r}_i - \overline{\mathbf{r}})(\mathbf{r}_i - \overline{\mathbf{r}})^T,$$

where

$$\overline{\mathbf{r}} = \begin{bmatrix} \overline{r}_1 \\ \overline{r}_2 \\ \vdots \\ \overline{r}_m \end{bmatrix}.$$

Obtain the eigenvalues of $\widehat{\Sigma}$ for different m's, and then fit the function $f(m) = \sigma^2 + 4\eta^2 \sin^2\left(\frac{\pi}{2(m+1)}\right)$ to the data. It is usually sufficient for the largest m value to be around 20. The estimate for σ^2 is then $\widehat{f(\infty)}$. This is called the fitted DST method. Figure 5.4.1 illustrates how the fitted DST method works. The annualised volatility was assumed to be 30% (indicated by the horizontal line). When σ^2 is estimated by the smallest eigenvalue i.e.$\widehat{\lambda}_m$ of $\widehat{\Sigma}$, this estimator is called the DST estimator.

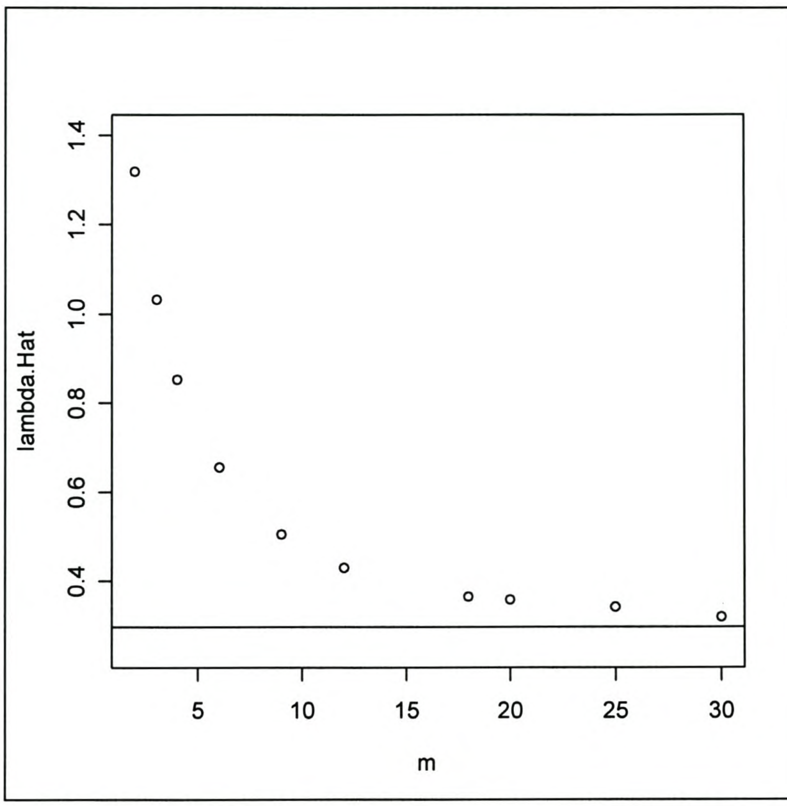

Figure 5.4.1: $\widehat{\lambda}_m$ as a function of m

Further, if we assume that $\mathbf{r} = \left[r_{t_n}, r_{t_{n-1}}, ..., r_{t_{n-m+1}}\right]^T \sim normal(\mathbf{0}; \Sigma)$, then we may obtain the Cramer-Rao bounds for $\hat{\sigma}^2$ and $\hat{\eta}^2$. These bounds are derived next and can be used to decide on the size of m.

Let

$$\mathbf{E} : m \times m = \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \vdots \\ \mathbf{e}_m^T \end{bmatrix}$$

$\Rightarrow$$\mathbf{Y}$= $E\mathbf{r} \sim normal(\mathbf{0}; \mathbf{E}\Sigma\mathbf{E}^T) = normal(\mathbf{0}; \Lambda)$, with $\Lambda : m \times m =$Diag$(\lambda_1, \lambda_2, ..., \lambda_m)$ because $\mathbf{E}$ is an orthogonal matrix (i-th column of $\mathbf{E}$ is the eigenvector assosiated with λ_i). Therefore

$$f_y(\mathbf{Y}) = (2\pi)^{-\frac{1}{2}m} |\Lambda|^{-\frac{1}{2}} eksp\left(-\frac{1}{2}\mathbf{Y}^T\Lambda^{-1}\mathbf{Y}\right).$$

Consider

$$\begin{aligned} \mathbf{Y}^T\Lambda^{-1}\mathbf{Y} &= [Y_1, Y_2, ...Y_m] \begin{bmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_m} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} = \left[\frac{Y_1}{\lambda_1}, \frac{Y_2}{\lambda_2}, ..., \frac{Y_m}{\lambda_m}\right] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} \\ &= \sum_{i=1}^{m} \left(\frac{Y_i^2}{\lambda_i}\right). \end{aligned}$$

$$\Rightarrow f_y(\mathbf{Y}) = (2\pi)^{-\frac{1}{2}m} |\pi_{i=1}^m \lambda_i|^{-\frac{1}{2}} eksp\left(\sum_{i=1}^{m}\left(\frac{Y_i^2}{\lambda_i}\right)\right)$$

or

$$\log_e f_y(\mathbf{Y}) = -\frac{1}{2}m\log_e(2\pi) - \frac{1}{2}\sum_{i=1}^{m}\log_e \lambda_i - \frac{1}{2}\sum_{i=1}^{m}\left(\frac{Y_i^2}{\lambda_i}\right).$$

Let

$$\boldsymbol{\theta} = \begin{bmatrix} \sigma^2 \\ \eta^2 \end{bmatrix}.$$

Then

$$\frac{\partial \lambda_i}{\partial \boldsymbol{\theta}} = \begin{bmatrix} 1 \\ 4\sin^2\left(\frac{\pi(m+1-i)}{2(m+1)}\right) \end{bmatrix}$$

and

$$\frac{\partial^2 \lambda_i}{\partial \theta_j \partial \theta_k} = 0 \quad j, k = 1, 2\,.$$

Thus

$$\frac{\partial \log_e f_y(\mathbf{Y})}{\partial \theta_j} = -\frac{1}{2}\sum_{i=1}^{m}\frac{1}{\lambda_i}\frac{\partial \lambda_i}{\partial \theta_j} + \frac{1}{2}\sum_{i=1}^{m}\frac{Y_i^2}{\lambda_i^2}\frac{\partial \lambda_i}{\partial \theta_j} = \frac{1}{2}\sum_{i=1}^{m}\left(\frac{Y_i^2}{\lambda_i^2} - \frac{1}{\lambda_i}\right)\frac{\partial \lambda_i}{\partial \theta_j}$$

and

$$-\frac{\partial^2 \log_e f_y(\mathbf{Y})}{\partial \theta_j \partial \theta_k} = -\frac{1}{2}\sum_{i=1}^{m}\frac{1}{\lambda_i^2}\frac{\partial \lambda_i}{\partial \theta_j}\frac{\partial \lambda_i}{\partial \theta_k} + \sum_{i=1}^{m}\frac{Y_i^2}{\lambda_i^3}\frac{\partial \lambda_i}{\partial \theta_j}\frac{\partial \lambda_i}{\partial \theta_k} = \sum_{i=1}^{m}\left(\frac{Y_i^2}{\lambda_i^3} - \frac{1}{2}\frac{1}{\lambda_i^2}\right)\frac{\partial \lambda_i}{\partial \theta_j}\frac{\partial \lambda_i}{\partial \theta_k}.$$

To obtain the Fisher Information matrix, remember that

$$I_{jk} = -E\left(\frac{\partial^2 \log_e f_y(\mathbf{Y})}{\partial \theta_j \partial \theta_k}\right),$$

with

$$\mathbf{Y} \sim normal(\mathbf{0}; \Lambda),$$

and thus

$$E\left(Y_i^2\right) = \lambda_i,$$

leading to

$$I_{jk} = \sum_{i=1}^{m}\left(\frac{\lambda_i}{\lambda_i^3} - \frac{1}{2}\frac{1}{\lambda_i^2}\right)\frac{\partial \lambda_i}{\partial \theta_j}\frac{\partial \lambda_i}{\partial \theta_k} = \frac{1}{2}\sum_{i=1}^{m}\frac{1}{\lambda_i^2}\frac{\partial \lambda_i}{\partial \theta_j}\frac{\partial \lambda_i}{\partial \theta_k}.$$

Specifically,

$$I_{11} = \frac{1}{2}\sum_{i=1}^{m}\frac{1}{\lambda_i^2},$$

$$I_{12} = I_{21} = 2\sum_{i=1}^{m}\frac{1}{\lambda_i^2}\sin^2\left(\frac{\pi(m+1-i)}{2(m+1)}\right)$$

and

$$I_{22} = 8\sum_{i=1}^{m}\frac{1}{\lambda_i^2}\sin^4\left(\frac{\pi(m+1-i)}{2(m+1)}\right).$$

Then the Cramer-Rao bounds for $\hat{\sigma}^2$ and $\hat{\eta}^2$ follow respectively as:

$$Var\left(\hat{\sigma}^2\right) \geq \frac{I_{22}}{I_{11}I_{22} - I_{12}^2}$$

and

$$Var\left(\hat{\eta}^2\right) \geq \frac{I_{11}}{I_{11}I_{22} - I_{12}^2}\,.$$

These bounds may be used to determine the m, such that the lower bounds of the variance of $\hat{\sigma}^2$ is some desired amount. Using simulations, one can approximate the variance of the DST measurement and compare it to the Cramer-Rao bound of $\hat{\sigma}^2$ determining the m, such that the difference between the variance of the measurement and the Cramer-Rao bound is small. How small the difference must be is up to the individual to decide.

6

Return and ex-ante volatility models

6.1 Introduction

As mentioned in Chapter 4, before reliable ex-post measurements had been formulated, reliable ex-ante volatility models could not be built because all the stylised facts were not yet known. Volatility forecasts use ex-post volatility estimates, and these estimates were either noisy or based on some or other model as discussed in Chapters 2 and 3. In Chapters 4 and 5 ex-post volatility measurements that are non-parametric and relatively error-free were discussed. These volatility measurements can be used to catch the most important stylised facts of the unobservable volatility process. These facts can then be used to build models for forecasting purposes using the volatility measurements mentioned. Furthermore, volatility forecasts use ex-post volatility estimates and the better the estimates, the better the forecasts should be. Return and volatility models used for forecasting purposes are the focus of attention in this chapter. Before considering the models, the stylised facts of returns and volatility need to be looked at again.

Stylised facts of returns and volatility:

i) Fat tails of high frequency returns (kurtosis >3), and the kurtosis tends to decrease as the time interval increases. This is called the cross-over effect.
ii) High frequency returns are autocorrelated.
iii) The volatility measurements are autocorrelated for up to at least a month (volatility clustering).

iv) Returns have the multifractal property: $\mathrm{E}(|p_{t+\Delta t} - p_t|^q) \sim (\Delta t)^{q.H(q)}$ where $H(q)$ is the Hurst exponent.

v) Volatiliy cascade characteristic: long-term views have a marked influence on the short-term views. Volatility over longer intervals has a larger influence on shorter intervals than conversely. (This is more important economically than mathematically). Volatility over shorter intervals can be written as a function of volatility over longer time intervals.

vi) Leverage effect: volatility tends to increase when prices are falling, and vice versa.

vii) $\frac{r_t}{\sigma_t}$ is normally distributed.

viii) σ_t is lognormally distributed.

In Section 6.2 a benchmark model is given that can be used when testing how good another model is, in Section 6.3 the forecasting volatility model suggested in Corsi (2003) is given and in Section 6.4 a summary is given.

6.2 The simplest possible model

The following return model may be used as a benchmark model:

$$dp_t = \mu dt + \sigma dB_t \ ,$$

where $r_t = p_t - p_0$ and B_t denotes a standard Brownian motion , i.e.

$$B_t - B_s \sim normal(0; t - s) \ ,$$

where B_t has stationary independent increments, with $B_o = 0$. Here we have:

$$r_t = \int_0^t dp_t = \int_0^t \mu dt + \int_0^t \sigma dB_t = \mu t + \sigma B_t \ , \tag{6.2.1}$$

with standard notation $\mu t \equiv A_t$, $\sigma B_t \equiv M_t^C$, and $M_t^J = 0$.

Suppose the realised sample is

$$r_{t_1,t_1-h/n}, r_{t_2,t_2-h/n}, \ldots, r_{t_n,t_n-h/n} \ ,$$

with $t_i - t_{i-1} = h/n$, and $t_i \in [0, t]$ for $i = 1, 2, \ldots, n$.

It follows from (6.2.1) that

$$r_t \sim normal(\mu t; \sigma^2 t).$$

Therefore the maximum likelihood estimate for μ is the sample mean of the scaled returns. That is

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{n}{h}r_{t_i,t_i-h/n} = \frac{r_{t,t-h}}{h}$$

and

$$E\left(\hat{\mu}_n\right) = E\left(\frac{r_{t,t-h}}{h}\right) = \frac{\mu h}{h} = \mu\,.$$

$\hat{\mu}_n$ is an unbiased estimate for μ, but is not consistent since

$$Var\left(\hat{\mu}_n\right) = Var\left(\frac{r_{t,t-h}}{h}\right) = \frac{\sigma^2 h}{h^2} = \frac{\sigma^2}{h}\,.$$

An estimator for σ^2 is

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{n}{h}r^2_{t_i,t_i-h/n}$$

and

$$\begin{aligned} E\left(\hat{\sigma}_n^2\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\frac{n}{h}r^2_{t_i,t_i-h/n}\right) \\ &= \frac{1}{h}\sum_{i=1}^{n}E(r^2_{t_i,t_i-h/n})\,, \end{aligned}$$

but

$$E(r^2_{t_i,t_i-h/n}) = Var(r_{t_i,t_i-h/n}) + (E(r_{t_i,t_i-h/n}))^2 = \sigma^2\frac{h}{n} + \mu^2\left(\frac{h}{n}\right)^2\,.$$

This implies that

$$E\left(\hat{\sigma}_n^2\right) = \sigma^2 + \mu^2\frac{h}{n}\,,$$

which is biased, so an unbiased estimator for σ^2 is given by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{n}{h}r^2_{t_i,t_i-h/n} - \mu^2\frac{h}{n}. \tag{6.2.2}$$

This unbiased estimator is a consistent estimator. To see this consider:

$$
\begin{aligned}
Var\left(\hat{\sigma}_n^2 - \mu^2\frac{h}{n}\right) &= Var\left(\hat{\sigma}_n^2\right) \\
&= Var\left(\frac{1}{n}\sum_{i=1}^{n}\frac{n}{h}r^2_{t_i,t_i-h/n}\right) \\
&= \frac{1}{h^2}\sum_{i=1}^{n}Var(r^2_{t_i,t_i-h/n}) \\
&= \frac{1}{h^2}\sum_{i=1}^{n}\{E(r^4_{t_i,t_i-h/n}) - (E(r^2_{t_i,t_i-h/n}))^2\} \\
&= \frac{1}{h^2}\sum_{i=1}^{n}\{(\mu\frac{h}{n})^4 + 6\sigma^2\frac{h}{n}(\mu\frac{h}{n})^2 + 3(\sigma^2\frac{h}{n})^2 - \\
&\quad ((\sigma^2\frac{h}{n})^2 + (\mu\frac{h}{n})^4 + 2\sigma^2\mu^2(\frac{h}{n})^3)\} \\
&= \frac{1}{h^2}\sum_{i=1}^{n}\{4\sigma^2\mu^2(\frac{h}{n})^3 + 2\sigma^4(\frac{h}{n})^2\} \\
&= 4\sigma^2\mu^2hn^{-2} + 2\sigma^4n^{-1}
\end{aligned}
$$

$$
\Rightarrow \lim_{n\to\infty} Var\left(\hat{\sigma}_n^2\right) = 0 \,.
$$

Suppose we are at time t, then the forecasted volatility between time t and $t+s$ is simply

$$
s \bullet \left(\frac{1}{n}\sum_{i=1}^{n}\frac{n}{h}r^2_{t_i,t_i-h/n} - \mu^2\frac{h}{n}\right).
$$

When deciding on a volatility model for forcasting purposes, the model in question needs to outperform model (6.2.2).

6.3 The suggested model for forecasting purposes

In Anderson, Bollerslev and Diebold (2003), and Corsi (2003) a return model and a volatility model, respectively, are suggested. The results given in these two papers are thoroughly derived, explained and discussed in the rest of this chapter.

6.3.1 The return model

The return model, of Anderson et al (2003), is:

$$dp_t = (\mu_t + \lambda\mu_\kappa)dt + \sigma_t dB_t + \kappa_t dN_t - \lambda\mu_\kappa dt \,, \tag{6.3.1.1}$$

where B_t is a standard Brownian motion, and N_t is a pure jump process, with $\Delta N_t = 1$, if jump at time t or 0 otherwise. The jump intensity is at a rate λ per time unit, while κ_t denotes the size of the jump at time t, with $\mathrm{E}(\kappa_t) = \mu_\kappa$, and $\mathrm{Var}(\kappa_t) = \sigma_\kappa^2$.

Here we have:

$$r_t = \int_0^t \mu_s ds + \lambda\mu_\kappa t + \int_0^t \sigma_s dB_s + \sum_{0\le t_s\le t} \kappa_{t_s}\Delta N_{t_s} - \lambda\mu_\kappa t,$$

with standard notation $\int_0^t \mu_s ds + \lambda\mu_\kappa t \equiv A_t$, $\int_0^t \sigma_s dB_s \equiv M_t^C$ and $\sum_{0\le t_s\le t} \kappa_{t_s}\Delta N_{t_s} - \lambda\mu_\kappa t \equiv M_t^J$. The $\lambda\mu_\kappa t$ term was added so that the martingale properties are satisfied.

To show some interesting characteristics, consider a simplification of Model (6.3.1.1):

$$dp_t = \mu_t dt + \sigma_t dB_t$$

$$\Rightarrow r_t = \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s \,.$$

Suppose we know the values of $\{\mu_s\}$ and $\{\sigma_s\}$, with $s \in [0,t]$. It then follows straightforwardly that $r_t \mid \{\mu_s\}\ \{\sigma_s\}$ (see Section 3.2 for notation) will be normally distributed with

$$E\left(r_t|\{\mu_s\}\{\sigma_s\}\right) = \int_0^t \mu_s ds$$

because

$$E\left(\int_0^t \sigma_s dB_s\right) = 0$$

and

$$\begin{aligned}
Var\left(r_t \mid \{\mu_s\}\{\sigma_s\}\right) &= Var\left(\int_0^t \mu_s ds\right) + Var\left(\int_0^t \sigma_s dB_s\right) \\
&= E\left(\int_0^t \sigma_s dB_s\right)^2 \\
&= E\left(\int_0^t \sigma_s^2 \left(dB_s\right)^2 + \int_{0s\neq r}^t \int_0^t \sigma_s \sigma_r dB_s dB_r\right), \quad \left(dB_s\right)^2 = ds \\
&= E\left(\int_0^t \sigma_s^2 ds\right) + 0, \text{ because Brownian motions has independent increments} \\
&= \int_0^t \sigma_s^2 ds \,.
\end{aligned}$$

So

$$r_t \mid \{\mu_s\}\{\sigma_s\} \sim normal\left(\int_0^t \mu_s ds \,;\; \int_0^t \sigma_s^2 ds\right).$$

If we let σ_t be a random variable, it follows that the unconditional distribution of r_t will be a fat-tailed distribution. [See Figure (6.3.1.1).] The returns were simulated using S-plus, with σ_t a lognormal random variable. This is the distribution often encountered in practice. It can be seen that the resulting distribution of r_t has a much fatter tail than the normal distribution. For the rigorous proof of this result see [Jarrow, 1998].

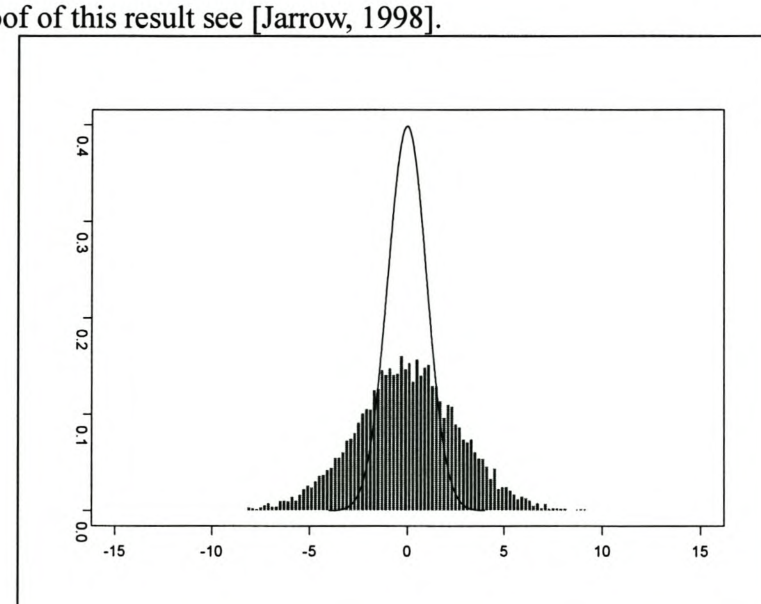

Figure 6.3.1.1 Normal vs fat tail distribution

Let us now look at how $[r,r]_t - [r,r]_{t-h}$ can be used to estimate the volatility for model (6.3.1.1) given a certain volatility model. We first need the following result:

$$\begin{aligned}
\left[M^C, M^C\right]_t &= (M_t^C)^2 - 2\int_0^t M_{s-}^C dM_s^C \\
&= \left(\int_0^t \sigma_s dB_s\right)^2 - 2\int_0^t \underbrace{\int_0^s \sigma_r dB_r}_{M_{s-}^C} dM_s^C\ , \\
\text{with } dM_s^C &= \int_{s^-}^s \sigma_s dB_s = \sigma_s\left(B_s - B_{s^-}\right) = \sigma_s dB_s\ .
\end{aligned}$$

Hence

$$\begin{aligned}
\left[M^C, M^C\right]_t &= \int_0^t \sigma_s^2 \left(dB_s\right)^2 + \underset{s\neq r}{\int_0^t \int_0^t} \sigma_s \sigma_r dB_s dB_r - 2 \underset{r<s}{\int_0^t \int_0^s} \sigma_r dB_r \sigma_s dB_s \\
&= \int_0^t \sigma_s^2 ds + \underset{s\neq r}{\int_0^t \int_0^t} \sigma_s \sigma_r dB_s dB_r - \underset{s\neq r}{\int_0^t \int_0^t} \sigma_r \sigma_s dB_r dB_s \\
&= \int_0^t \sigma_s^2 ds\ .
\end{aligned}$$

From (4.4.3) we have that

$$\begin{aligned}
[r,r]_t - [r,r]_{t-h} &= \left[M^C, M^C\right]_t + \sum_{0\leq t_s \leq t} \kappa_{t_s}^2 \Delta N_{t_s}^2 - \left(\left[M^C, M^C\right]_{t-h} + \sum_{0\leq t_s\leq t-h} \kappa_{t_s}^2 \Delta N_{t_s}^2\right) \\
&= \int_{t-h}^t \sigma_s^2 ds + \sum_{t-h\leq t_s\leq t} \kappa_{t_s}^2 \Delta N_{t_s}^2\ . \qquad (6.3.1.2)
\end{aligned}$$

Remembering that in *Barndorff-Nielsen and Shepard (2003)* it was shown that

$$\left[M^C, M^C\right]_t = \lim_{n\to\infty} \frac{\pi}{2} \sum_{k=1}^{n} \mid r_{t_k,t_{k-1}} \mid\mid r_{t_{k-1},t_{k-2}} \mid\ , \text{ where } t_k \in [0,t] \text{ for all k.}$$

Hence

$$\begin{aligned}\sum_{t-h\leq t_s\leq t}\kappa_{t_s}^2\Delta N_{t_s}^2 &= [r,r]_t-[r,r]_{t-h}-\int_{t-h}^{t}\sigma_s^2 ds\\ &= [r,r]_t-[r,r]_{t-h}-\left[M^C,M^C\right]_t\\ &\approx \sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2-\frac{\pi}{2}\sum_{k=1}^{n}\mid r_{t_k,t_{k-1}}\mid\mid r_{t_{k-1},t_{k-2}}\mid\text{, with } t_k\in[0,t].\end{aligned}$$

So $\sum_{0\leq t_s\leq t}\kappa_{t_s}^2\Delta N_{t_s}^2$ can be approximated by:

$$RJ_{t,0}=RV_{t,0}-RB_{t,0}=\sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2-\frac{\pi}{2}\sum_{k=1}^{n}\mid r_{t_k,t_{k-1}}\mid\mid r_{t_{k-1},t_{k-2}}\mid\text{, with } t_k\in[0,t] \tag{6.3.1.3}$$

if this term is positive, otherwise approximate $\sum_{0\leq t_s\leq t}\kappa_{t_s}^2\Delta N_{t_s}^2$ by zero.

6.3.2 The volatility model

Let the unit time be in days, and define the following:

$$\begin{aligned}RV_{t+h,t} &= \frac{1}{h}\left(\sum_{i=1}^{h}RV_{t+i,t-1+i}\right)\\ &= \frac{1}{h}\left(\sum_{i=1}^{h}\sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2\right)\\ \text{where } t_0 &= t\text{ and }t_n=t+h\end{aligned}$$

and

$$\begin{aligned}RJ_{t+h,t} &= \frac{1}{h}\left(\sum_{i=1}^{h}RJ_{t+i,t-1+i}\right)\\ &= \frac{1}{h}\left(\sum_{i=1}^{h}(6.3.1.3)\right)\\ \text{where } t_0 &= t\text{ and }t_n=t+h,\end{aligned}$$

i.e. $RV_{t+h,t}$ is the average estimated daily volatility between day t and $t+h$.

The volatility model that was suggested in Corsi (2003) is:

$$(RV_{t+1,t})^{\frac{1}{2}} = \beta_0 + \beta_D(RV_{t,t-1})^{\frac{1}{2}} + \beta_W(RV_{t,t-5})^{\frac{1}{2}} + \beta_M(RV_{t,t-22})^{\frac{1}{2}} + \beta_J(RJ_{t,t-1})^{\frac{1}{2}} + \varepsilon_{t+1}, \tag{6.3.2.1}$$

with $E\left(\varepsilon_{t+1}\right) = 0$

In model (6.3.2.1) $RV_{t+1,t}$ is tommorow's volatility, the volatilities on the rigth-hand side of (6.3.2.1) are all ex-post volatilities i.e. it can be approximated at time t and may be approximated by (4.3.1) or using the discrete sine transform technique. $RJ_{t,t-1}$ is approximated by (6.3.1.3). **It can be shown that model (6.3.2.1) caters for all the stylised facts mentioned in Section 6.1.**

To forecast over longer time- horizons the authors assumed that

$$E\left(RV_{t+h,t}\right)^{\frac{1}{2}} = E\left(RV_{t+1,t}\right)^{\frac{1}{2}} \ ,$$

so that:

$$\begin{aligned} (RV_{t+h,t})^{\frac{1}{2}} &= E\left(RV_{t+h,t}\right)^{\frac{1}{2}} + \varepsilon_{t+h,t} = E\left(RV_{t+1,t}\right)^{\frac{1}{2}} + \varepsilon_{t+h,t} \\ &= \beta_0 + \beta_D(RV_{t,t-1})^{\frac{1}{2}} + \beta_W(RV_{t,t-5})^{\frac{1}{2}} + \beta_M(RV_{t,t-22})^{\frac{1}{2}} + \\ &\quad \beta_J(RJ_{t,t-1})^{\frac{1}{2}} + \varepsilon_{t+h,t} \end{aligned} \tag{6.3.2.2}$$

and

$$\lim_{n\to\infty} RV_{t+h,t} = \frac{1}{h}\left(\sum_{i=1}^{h} \lim_{n\to\infty} RV_{t+i,t-1+i}\right) = \frac{1}{h}\sum_{i=1}^{h}(\lim_{n\to\infty}\sum_{k=1}^{n}[r_{t_{k_i},t_{k_i}-1}]^2)\ , \tag{6.3.2.3}$$

where $t_{k_i} \in [t-1+i, t+i]$ and $\varepsilon_{t+h,t} = \frac{1}{h}\sum_{i=1}^{h}\varepsilon_{t+i,t-1+i}$. Here again the term $RV_{t+h,t}$ in (6.3.2.2) is the future volatility and the terms on the right-hand side of (6.3.2.2) are past volatilities. The unknown parameters in (6.3.2.1) and (6.3.2.2) can be estimated by linear regression.

From (6.3.2.3) and $[r,r]_t - [r,r]_{t-h} = p\lim_{n\to\infty} \sum_{k=1}^{n} r^2_{t_k,t_{k-1}}$, it follows that

$$\lim_{n\to\infty} RV_{t+h,t} = \frac{1}{h}\sum_{i=1}^{h}([r,r]_{t+i} - [r,r]_{t-1+i}) \, .$$

Thus using (6.3.1.2) it follows that

$$\begin{aligned}
\lim_{n\to\infty} E(RV_{t+h,t}) &= \frac{1}{h}\sum_{i=1}^{h} E([r,r]_{t+i} - [r,r]_{t-1+i}) \\
&= \frac{1}{h}\sum_{i=1}^{h}\{\int_{t-1+i}^{t+i} \sigma_s^2 ds + \lambda(\sigma_\kappa^2 + \mu_\kappa^2)\} \\
&= \frac{1}{h}\int_{t}^{t+h} \sigma_s^2 ds + \lambda(\sigma_\kappa^2 + \mu_\kappa^2)
\end{aligned}$$

and also

$$\lim_{n\to\infty} E(RV_{t+1,t}) = \int_{t}^{t+1} \sigma_s^2 ds + \lambda(\sigma_\kappa^2 + \mu_\kappa^2) \, .$$

It follows that they actually assume that

$$\frac{1}{h}\int_{t}^{t+h} \sigma_s^2 ds = \int_{t}^{t+1} \sigma_s^2 ds$$

which is the best they can do given the information up to time t.

$RV_{t,0}$ can be measured using either the realised volatility measurement or the DST measurement. If one looks at the model it is clear that the error terms $\{\varepsilon_{t+i}\}$ will be serially correlated up to at least order $h-1$. When estimating the standard errors of the coefficients of the model, we need to use the Newey-West correction to ensure that the covariance matrix is semi- positive definitive. [See Hamilton, 1994 for more on this topic.]

6.4 Summary

The most important results shown in this chapter were, that if we assume the return model to be

$$dp_t = (\mu_t + \lambda\mu_\kappa)dt + \sigma_t dB_t + \kappa_t dN_t - \lambda\mu_\kappa dt$$

then

$$[r,r]_t - [r,r]_{t-h} = \int_{t-h}^{t} \sigma_s^2 ds + \sum_{t-h \le t_s \le t} \kappa_{t_s}^2 \Delta N_{t_s}^2,$$

$$\int_{t-h}^{t} \sigma_s^2 ds \approx \frac{\pi}{2} \sum_{k=1}^{n} \mid r_{t_k,t_{k-1}} \mid\mid r_{t_{k-1},t_{k-2}} \mid$$

and

$$\sum_{t-h \le t_s \le t} \kappa_{t_s}^2 \Delta N_{t_s}^2, \approx \sum_{k=1}^{n} [r_{t_k,t_{k-1}}]^2 - \frac{\pi}{2} \sum_{k=1}^{n} \mid r_{t_k,t_{k-1}} \mid\mid r_{t_{k-1},t_{k-2}} \mid \ .$$

A volatility model, that caters for all the stylised facts, is

$$(RV_{t+h,t})^{\frac{1}{2}} = \beta_0 + \beta_D (RV_t)^{\frac{1}{2}} + \beta_W (RV_{t,t-5})^{\frac{1}{2}} + \beta_M (RV_{t,t-22})^{\frac{1}{2}} + \beta_J (RJ_{t,t-1})^{\frac{1}{2}} + \varepsilon_{t+h,t} \ ,$$

where the term $RV_{t+h,t}$ is the future volatility and the terms on the right-hand side are past volatilities. The unknown parameters can be estimated by linear regression.

7

A realised volatility measurement using quadratic variation handling microstructure effects

7.1 Introduction

In this chapter a volatility measurement is given that overcomes the problems experienced by the realised volatility and the DST volatility. The shortcomings of these measurements are discussed in detail first.

The problem of the realised volatility measurement is that as the sample frequency increases, market microstructure effects influence the assumption that returns are independently distributed. Bid-ask spread and price discreteness are the main factors of microstructure effects. [See Cohen et al ,1981, Glosten, 1994, Roll, 1984 and Glottlieb and Kalay, 1985 for more on this topic.] Owing to these microstructure effects, it no longer holds that the variance of the sum is the sum of the variances of the returns. Corsi and Cursi (2003) demonstrated how to measure volatility in the presence of non-zero autocorrelation of returns based on the Discrete Sine Transform (*DST*) approach. This method is appropriate if the instantaneous volatility is constant, but in practice this is not the case, as has already been pointed out. However, this measurement still gives good estimates of annualised volatility should the instantaneous volatility not change too much because a weighted average instantaneous volatility over the time horizon is used. In practice this method cannot be employed to obtain volatility estimates over short-time intervals, i.e. daily or weekly estimates. Although this measurement may yield good annualised volatility estimates, for any given year it will give constant daily volatility estimates, which is not satisfatory. One way of overcoming this problem is to sample returns over a short enough time interval such that the instantaneous volatility over that period is approximately constant. In reality though, too short time intervals are then used, resulting in noisy volatility estimates. There is still a need for a volatility measurement that can manage microstructure effects and varying instantaneous volatility. The realised volatility measurement based on quadratic variation martingale theory handles varying instantaneous volatility effectively, but fails to deal with the microstructure effects. Similarly, the volatility measurement, based on the DST approach, manages the microstructure effects efficiently, but breaks down under a changing instantaneous volatility environment.

A modified realised volatility measurement is subsequently defined (see Section 7.2), which we term *microstructure realised volatility*. It attempts to address some of the problems of realised volatility based on the quadratic variation and the DST approach. This method is also based on quadratic variation theory, but the underlying return model is more realistic and incorporates microstructure effects. This measurement can therefore handle both microstructure effects and non-constant volatility. Only the derivation for first-order autocorrelation is given, but to extend the model for other autocorrelation lags is straightforward.

7.2 A modified realised volatility measurement

The model in (4.1.1) is not realistic when the sampling frequency is very small because it assumes that returns are independent. A model that can handle dependency of returns of one or more lags is needed.

A more applicable return model is:

$$r_t = A_t + M_t + \phi M_{t-s}, \tag{7.2.1}$$

where s is the time between consecutive trades and where A_t denotes the predictable process, M_t and M_{t-s} are unpredictable processes (martingales) as before and with $0 \leq \phi \leq 1$.

Model (7.2.1) assumes first order autocorrelation of returns. Consider:

$$
\begin{aligned}
& cov\{(r_t - r_{t-s})(r_{t-s} - r_{t-2s})|F_{t-h}\} \\
= \; & E\{(r_t - r_{t-s})(r_{t-s} - r_{t-2s}) \mid F_{t-h}\} \\
& -E\{(r_t - r_{t-s})|F_{t-h}\}E\{(r_{t-s} - r_{t-2s})|F_{t-h}\} \\
= \; & E\{(A_t - A_{t-s} + M_t - M_{t-s} + \phi\,(M_{t-s} - M_{t-2s}))\cdot \\
& (A_{t-s} - A_{t-2s} + M_{t-s} - M_{t-2s} + \phi\,(M_{t-2s} - M_{t-3s}))|F_{t-h}\} \\
& -E\,(A_t - A_{t-s}|F_{t-h})\,E\,(A_{t-s} - A_{t-2s}|F_{t-h}) \\
= \; & E\{(A_t - A_{t-s} + M_t - M_{t-s} + \phi\,(M_{t-s} - M_{t-2s}))\cdot \\
& (A_{t-s} - A_{t-2s} + M_{t-s} - M_{t-2s} + \phi\,(M_{t-2s} - M_{t-3s}))|F_{t-h}\} \\
& -E\,(A_t - A_{t-s}|F_{t-h})\,E\,(A_{t-s} - A_{t-2s}|F_{t-h}) \\
= \; & E\{(A_t - A_{t-s})\,A_{t-s} - A_{t-2s} \mid F_{t-h}\} \\
& +\phi E\{(M_{t-s} - M_{t-2s})^2|F_{t-h}\} \\
& -E\,(A_t - A_{t-s}|F_{t-h})\,E\,(A_{t-s} - A_{t-2s}|F_{t-h}) \\
= \; & \phi E\{(M_{t-s} - M_{t-2s})^2|F_{t-h}\} \\
& +E\,(A_t - A_{t-s} \mid F_{t-h})\,E\,(A_{t-s} - A_{t-2s}|F_{t-h}) \\
& -E\,(A_t - A_{t-s}|F_{t-h})\,E\,(A_{t-s} - A_{t-2s}|F_{t-h})
\end{aligned}
$$

because the $\{A_t\}$ are independent in a time window of length s.

Therefore

$$cov\{(r_t - r_{t-s})(r_{t-s} - r_{t-2s})|F_{t-h}\} = \phi E\{(M_{t-s} - M_{t-2s})^2 \mid F_{t-h}\}$$

and if we assume M_t to be a Brownian motion, we have

$$cov\{(r_t - r_{t-s})(r_{t-s} - r_{t-2s})|F_{t-h}\} = \phi s.$$

The same path as was taken for Model (4.1.1), needs to be taken for Model (7.2.1) in order to procure an ex-post measurement for the volatility of Model (7.2.1). In the following theorem the variance of the return process is expressed in terms of M_t.

Theorem 7.2.1

$$var(r_t|F_{t-h}) = E(M_t^2|F_{t-h}) - M_{t-h}^2 + \phi\,(\phi+2)\,(E(M_t^2|F_{t-h}) - M_{t-h}^2)$$

over short-time horizons.

Proof

$$\begin{aligned}
var(r_t|F_{t-h}) &= E\{(r_t - E(r_t|F_{t-h}))^2|F_{t-h}\} \\
&= E\{(A_t + M_t + \phi M_{t-s} - E(A_t + M_t + \phi M_{t-s}|F_{t-h}))^2|F_{t-h}\} \\
&= E\{(A_t + M_t + \phi M_{t-s} - E(A_t|F_{t-h}) - M_{t-h} - \phi M_{t-h})^2|F_{t-h}\} \\
&= E\{((A_t - E(A_t|F_{t-h}) + (M_t - M_{t-h}) + \phi(M_{t-s} - M_{t-h}))^2|F_{t-h}\} \\
&= E\{(M_t - M_{t-h})^2\,|F_{t-h}\} + \phi^2 E\{(M_{t-s} - M_{t-h})^2\,|F_{t-h}\} \\
&\quad +2\phi E\{(M_t - M_{t-h})\,(M_{t-s} - M_{t-h})\,|F_{t-h}\} + var(A_t|F_{t-h}) \\
&\quad +2\phi cov(A_t M_t|F_{t-h}) + 2\phi cov(A_t M_{t-s}|F_{t-h}) \\
&= E(M_t^2|F_{t-h}) - M_{t-h}^2 + \phi^2(E(M_t^2|F_{t-h}) - M_{t-h}^2) \\
&\quad +2\phi E\{(M_t - M_{t-s} + M_{t-s} - M_{t-h})\,(M_{t-s} - M_{t-h})\,|F_{t-h}\} \\
&\quad +var(A_t|F_{t-h}) + 2\phi cov(A_t M_t|F_{t-h}) + 2\phi cov(A_t M_{t-s}|F_{t-h}) \\
&= E(M_t^2|F_{t-h}) - M_{t-h}^2 + \phi^2(E(M_t^2|F_{t-h}) - M_{t-h}^2) \\
&\quad +2\phi E\{(M_{t-s} - M_{t-h})^2\,|F_{t-h}\} + var(A_t|F_{t-h}) \\
&\quad +2\phi cov(A_t M_t|F_{t-h}) + 2\phi cov(A_t M_{t-s}|F_{t-h}) \\
&= E(M_t^2|F_{t-h}) - M_{t-h}^2 + \phi\,(\phi+2)\,(E(M_t^2|F_{t-h}) - M_{t-h}^2) \qquad (7.2.2) \\
&\quad +var(A_t|F_{t-h}) + 2\phi cov(A_t M_t|F_{t-h}) + 2\phi cov(A_t M_{t-s}|F_{t-h})\,.
\end{aligned}$$

The last three terms in (7.2.2) are usually negligible and have an influence only over long-time horizons. This proves the theorem.

The quadratic variation process of Model (7.2.1) is now derived:

$$
\begin{aligned}
[r,r]_t &= r_t^2 - 2\int_0^t r_{q-}dr_q \\
&= (A_t + M_t + \phi M_{t-s})^2 - 2\int_0^t (A_{q-} + M_{q-} + \phi M_{q-s-})dr_q \\
&= A_t^2 + M_t^2 + \phi M_{t-s}^2 + 2A_tM_t + 2\phi A_tM_{t-s} + 2\phi M_tM_{t-s} \\
&\quad -2\int_0^t A_{q-}dA_q - 2\int_0^t M_{q-}dA_q - 2\phi\int_s^t M_{q-s-}dA_q - 2\int_0^t A_{q-}dM_q \\
&\quad -2\int_0^t M_{q-}dM_q - 2\phi\int_s^t M_{q-s-}dM_q - 2\phi\int_0^{t-s} A_{q-}dM_{q-s} \\
&\quad -2\phi\int_0^{t-s} M_{q-}dM_{q-s} - 2\phi^2\int_0^t M_{q-s-}dM_{q-s} \\
&= [M,M]_t + \phi^2 [M,M]_{t-s} \\
&\quad +2\phi\left(M_tM_{t-s} - \int_s^t M_{q-s-}dM_q - \int_0^{t-s} M_{q-}dM_{q-s}\right) \qquad (7.2.3)
\end{aligned}
$$

because the terms involving $\{A_t\}$ are zero.

Definition 7.2.1

Define

$$
[M, M_{-s}]_t = M_tM_{t-s} - \int_s^t M_{q-s-}dM_q - \int_0^{t-s} M_{q-}dM_{q-s} .
$$

With this notation:

$$
[r,r]_t = [M,M]_t + \phi^2 [M,M]_{t-s} + 2\phi [M, M_{-s}]_t \ . \qquad (7.2.4)
$$

Definition 7.2.1 has a significant implication for ex-post volatility measurements if the return process is given by Model (7.2.1). Consider:

$$
\begin{aligned}
[M, M_{-s}]_t &= M_t M_{t-s} - \int_s^t M_{q-s-} dM_q - \int_0^{t-s} M_{q-} dM_{q-s} \\
&= M_t M_{t-s} - p \lim_{n\to\infty} \{\sum_{k=s}^{n} M_{t_{k-s-1}} [M_{t_k} - M_{t_{k-1}}]\} \\
&\quad - p \lim_{n\to\infty} \{\sum_{k=s}^{n} M_{t_{k-1}} [M_{t_{k-s}} - M_{t_{k-s-1}}]\} \\
&= M_t M_{t-s} - p \lim_{n\to\infty} \{\sum_{k=s}^{n} (M_{t_{k-s-1}} [M_{t_k} - M_{t_{k-1}}] + M_{t_{k-1}} [M_{t_{k-s}} - M_{t_{k-s-1}}]\} \\
&= p \lim_{n\to\infty} \{\sum_{k=s}^{n} (M_{t_k} M_{t_{k-s}} - M_{t_{k-1}} M_{t_{k-s-1}})\} \\
&\quad - p \lim_{n\to\infty} \{\sum_{k=s}^{n} (M_{t_{k-s-1}} [M_{t_k} - M_{t_{k-1}}] + M_{t_{k-1}} [M_{t_{k-s}} - M_{t_{k-s-1}}]\} \\
&= p \lim_{n\to\infty} \{\sum_{k=s}^{n} (M_{t_k} M_{t_{k-s}} - M_{t_{k-1}} M_{t_{k-s-1}} - M_{t_{k-s-1}} [M_{t_k} - M_{t_{k-1}}] \\
&\quad - M_{t_{k-1}} [M_{t_{k-s}} - M_{t_{k-s-1}}])\} \\
&= p \lim_{n\to\infty} \{\sum_{k=s}^{n} (M_{t_k} M_{t_{k-s}} - M_{t_{k-1}} M_{t_{k-s-1}} - M_{t_{k-s-1}} [M_{t_k} - M_{t_{k-1}}] \\
&\quad - M_{t_{k-1}} [M_{t_{k-s}} - M_{t_{k-s-1}}])\} \\
&= p \lim_{n\to\infty} \{\sum_{k=s}^{n} (M_{t_k} M_{t_{k-s}} - M_{t_k} M_{t_{k-s-1}} - M_{t_{k-1}} M_{t_{k-s}} + M_{t_{k-1}} M_{t_{k-s-1}})\} \\
&= p \lim_{n\to\infty} \{\sum_{k=s}^{n} (M_{t_k} [M_{t_{k-s}} - M_{t_{k-s-1}}] - M_{t_{k-1}} [M_{t_{k-s}} + M_{t_{k-s-1}}])\} \\
&= p \lim_{n\to\infty} \{\sum_{k=s}^{n} (M_{t_k} - M_{t_{k-1}})(M_{t_{k-s}} - M_{t_{k-s-1}})\} \, .
\end{aligned}
$$

Therefore

$$[r, r_{-s}]_t \tag{7.2.5}$$

can be approximated by

$$\sum_{k=s}^{n} (r_{t_k, t_{k-1}})(r_{t_{k-1}, t_{k-2}}) \, ,$$

where n is made as large as possible.

We are going to show that the variance of model (7.2.1) can be written in terms of (7.2.4) and (7.2.5). Consider expression (7.2.5)

$$
\begin{aligned}
[r, r_{-s}]_t &= r_t r_{t-s} - \int_s^t r_{q-s-} dr_q - \int_0^{t-s} r_{q-} dr_{q-s} \\
&= (A_t + M_t + \phi M_{t-s})(A_{t-s} + M_{t-s} + \phi M_{t-2s}) \\
&\quad - \int_s^t (A_{q-s-} + M_{q-s-} + \phi M_{q-2s-}) dr_q - \int_0^{t-s} (A_{q-} + M_{q-} + \phi M_{q-s-}) dr_{q-s} \\
&= A_t A_{t-s} + A_t M_{t-s} + \phi A_t M_{t-2s} + M_t A_{t-s} + M_t M_{t-s} + \phi M_t M_{t-2s} \\
&\quad + \phi M_{t-s} A_{t-s} + \phi M_{t-s}^2 + \phi^2 M_{t-s} M_{t-2s} - \int_s^t A_{q-s-} dA_q - \int_s^t A_{q-s-} dM_q \\
&\quad - \phi \int_s^t A_{q-s-} dM_{q-s} - \int_s^t M_{q-s-} dA_q - \int_s^t M_{q-s-} dM_q \\
&\quad - \phi \int_s^t M_{q-s-} dM_{q-s} - \phi \int_s^t M_{q-2s-} dA_q - \phi \int_s^t M_{q-2s-} dM_q \\
&\quad - \phi^2 \int_s^t M_{q-2s-} dM_{q-s} - \int_s^t A_{q-} dA_{q-s} \\
&\quad - \int_s^t A_{q-} dM_{q-s} - \phi \int_s^t A_q dM_{q-2s} - \int_s^t M_{q-} dA_{q-s} \\
&\quad - \int_s^t M_{q-} dM_{q-s} - \phi \int_s^t M_q dM_{q-2s} - \phi \int_s^t M_{q-s-} dA_{q-s} \\
&\quad - \phi \int_s^t M_{q-s-} dM_{q-s} - \phi^2 \int_s^t M_{q-s} dM_{q-2s} \\
&= [M, M_{-s}]_t + \phi [M, M_{-2s}]_t + \phi [M, M]_{t-s} + \phi^2 [M, M_{-s}]_{t-s} \qquad (7.2.6)
\end{aligned}
$$

because the terms involving $\{A_t\}$ are zero.

Theorem 7.2.2

$$E\,[M, M_{-s}]_t = 0$$

Proof

$$\begin{aligned}
E([M, M_{-s}]_t - [M, M_{-s}]_{t-h}|F_{t-h}) &= E(M_t M_{t-s} - \int_s^t M_{q-s-} dM_q \\
&\quad - \int_0^{t-s} M_{q-} dM_{q-s} - M_{t-h} M_{t-s-h} \\
&\quad + \int_s^{t-h} M_{q-s-} dM_q + \int_0^{t-s-h} M_{q-} dM_{q-s}|F_{t-h}) \\
&= E(M_t M_{t-s}|F_{t-h}) - E(\int_{t-h}^t M_{q-s-} dM_q|F_{t-h}) \\
&\quad -E(\int_{t-s-h}^{t-s} M_{q-} dM_{q-s}|F_{t-h}) - M_{t-h} M_{t-s-h} \\
&= E((M_t - M_{t-s} + M_{t-s}) M_{t-s}|F_{t-h}) \\
&\quad -E(\int_{t-h}^t M_{q-s-} dM_q) \\
&\quad -E(\int_{t-s-h}^{t-s} (M_{t-s} - M_{t-s} + M_{q-}) dM_{q-s}|F_{t-h}) \\
&\quad -M_{t-h} M_{t-s-h} \\
&= 0 + E(M_{t-s}^2|F_{t-h}) - 0 \\
&\quad -E(\int_{t-s-h}^{t-s} (M_{t-s} dM_{q-s}|F_{t-h}) \\
&\quad +(E\int_{t-s-h}^{t-s} (M_{t-s} - M_{q-}) dM_{q-s}|F_{t-h}) \\
&\quad -M_{t-h} M_{t-s-h} \\
&= E(M_{t-s}^2|F_{t-h}) - E(M_{t-s}(M_{t-s} - M_{t-s-h})|F_{t-h}) \\
&\quad +0 - M_{t-h} M_{t-s-h} \\
&= E(M_{t-s}^2|F_{t-h}) - E(M_{t-s}^2|F_{t-h}) \\
&\quad +M_{t-h} M_{t-s-h} - M_{t-h} M_{t-s-h} \\
&= 0.
\end{aligned}$$

This completes the proof.

From (7.2.3), (7.2.6) and Theorem (7.2.2) the expected values of $[r, r]_t$ and $[r, r_{-s}]_t$ respectively, are

$$\begin{aligned} E\left([r,r]_t - [r,r]_{t-h}\,|F_{t-h}\right) &= E([M,M]_t + \phi^2\,[M,M]_{t-s} - [M,M]_{t-h} \\ &\quad +\phi^2\,[M,M]_{t-s-h}\,|F_{t-h}) \\ &= E\left(M_t^2|F_{t-h}\right) - M_{t-h}^2 \\ &\quad +\phi^2 E\left(M_{t-s}^2|F_{t-h}\right) - \phi^2 M_{t-s-h}^2 \end{aligned} \tag{7.2.7}$$

and

$$\begin{aligned} E\left([r,r_{-s}]_t - [r,r_{-s}]_{t-h+s}\,|F_{t-h}\right) &= E\left(\phi[M,M]_{t-s} - \phi[M,M]_{t-h}|F_{t-h}\right) \\ &= \phi E\left(M_{t-s}^2|F_{t-h}\right) - \phi M_{t-h}^2\,. \end{aligned} \tag{7.2.8}$$

Combining (7.2.7) and (7.2.8) it follows that:

$$\begin{aligned} & E\left([r,r]_t - [r,r]_{t-h} \mid F_{t-h}\right) + 2E\left([r,r_{-s}]_t - [r,r_{-s}]_{t-h+s} \mid F_{t-h}\right) \\ = \; & E\left(M_t^2 \mid F_{t-h}\right) - M_{t-h}^2 + \phi^2 E\left(M_{t-s}^2 \mid F_{t-h}\right) - \phi^2 M_{t-s-h}^2 \\ & +2\left(\phi E\left(M_{t-s}^2 \mid F_{t-h}\right) - \phi M_{t-h}^2\right) \\ = \; & E\left(M_t^2 \mid F_{t-h}\right) - M_{t-h}^2 + \phi\left(\phi+2\right)\left(E\left(M_{t-s}^2 \mid F_{t-h}\right) - M_{t-h}^2\right) \\ & +\phi^2\left(M_{t-h}^2 - M_{t-s-h}^2\right) \\ = \; & var(r_t \mid F_{t-h}) + \phi^2\left(M_{t-h}^2 - M_{t-s-h}^2\right)\,. \end{aligned}$$

The bias of this estimator is $\phi^2\left(M_{t-h}^2 - M_{t-s-h}^2\right)$. However, in practice ϕ^2 should be small, and $M_{t-h}^2 - M_{t-s-h}^2$ should be of the order of s. So

$$var(r_t \mid F_{t-h}) \approx E\left([r,r]_t - [r,r]_{t-h} \mid F_{t-h}\right) + 2E\left([r,r_{-s}]_t - [r,r_{-s}]_{t-h+s} \mid F_{t-h}\right)$$

and

$$\begin{aligned} [r,r]_t - [r,r]_{t-h} &\approx \sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2 \\ [r,r_{-s}]_t - [r,r_{-s}]_{t-h+s} &\approx \sum_{k=2}^{n}(r_{t_k,t_{k-1}})(r_{t_{k-1},t_{k-2}})\,. \end{aligned}$$

Remark 7.2

The variance of Model (7.2.1) can therefore be approximated by

$$\sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2 + 2\sum_{k=2}^{n}(r_{t_k,t_{k-1}})(r_{t_{k-1},t_{k-2}})\,.$$

The term $2\sum_{k=2}^{n}(r_{t_k,t_{k-1}})(r_{t_{k-1},t_{k-2}})$ is the difference between approximating the variance of returns for models (4.3.1) and (7.2.1).

If a lag m return model is used, i.e.

$$r_t = A_t + M_t + \phi_1 M_{t-s} + \phi_2 M_{t-2s} + ... + \phi_n M_{t-ms} \tag{7.2.9}$$

then

$$\sum_{k=1}^{n}[r_{t_k,t_{k-1}}]^2 + 2\sum_{k=2}^{n}(r_{t_k,t_{k-1}})(r_{t_{k-1},t_{k-2}}) + 2\sum_{k=3}^{n}(r_{t_k,t_{k-1}})(r_{t_{k-2},t_{k-3}})$$
$$+ ... + 2\sum_{k=m+1}^{n}(r_{t_k,t_{k-1}})(r_{t_{k-n},t_{k-(m+1)}})$$

is an ex-post measurement of volatility. The mathematics becomes rather unkempt, so the derivation of the variance of Model (7.2.9) is not given in this thesis.

7.3 Summary

In this chapter a volatility measurement was given that overcomes the problems experienced by the realised volatility and the DST volatility. The problem of the realised volatility measurement is that as the sample frequency increases, market microstructure effects influence the assumption that returns are independently distributed. Owing to these microstructure effects, it no longer holds that the variance of the sum is the sum of the variances of the returns.

The problem with the DST measurement is that it implicitly assumes constant instantaneous volatility, but in practice this is not the case. Although this measurement may still yield good annualised volatility estimates, for any given year it will give constant daily volatility estimates, which is not satisfatory.

A more applicable return model is

$$r_t = A_t + M_t + \phi M_{t-s}.$$

The variance of this return process can be approximated by

$$\sum_{k=1}^{n} [r_{t_k, t_{k-1}}]^2 + 2 \sum_{k=2}^{n} (r_{t_k, t_{k-1}})(r_{t_{k-1}, t_{k-2}}) .$$

Higher sample frequencies can be used for the *microstructure realised volatility* measurement, than for the *realised volatility* measurement, making it a better measurement. The *microstructure realised volatility* measurement can also easily be modified, to take higher than first order autocorrelation of returns into consideration.

8

Practical simulations and results

The same method to model the influence of microstructure effects as described in Hasbrouck (1993,1996), and as used in Corsi and Cursi (2003) is used here. 72000 returns of length 5 minutes are simulated a 100 times, with different characteristics in each simulation. We are to compare the microstructure realised volatility (micr.RV) with the realised volatility (RV), the 30 minutes realised volatility (every 6th return is to be taken) and the DST- measurement. In the tables the mean, the mean squared error (MSE) and the variance (Var) of the 4 measurements are given. In determining the DST measurement m is taken as 40, and the bias is subtracted. The DST measurement is therefore unbiased.

Simulation 1: Table 8.1

Table 8.1: Constant instantaneous volatility and no autocorrelation			
Annualised volatility: 29.5161 %			
	mean	MSE	Var
micr.RV	29.5350	0.0175	0.0173
DST	28.5715	0.9107	0.0186
RV	29.5233	0.0044	0.0044
RV.30min	29.5356	0.0335	0.0334

The returns are independent with constant instantaneous volatility. From Table 8.1 it is clear that all estimators performs well. The means are close to 29.5161, the mean squared errors and variances are small with the realised volatility measurement that outperforms the other in this regard.

Simulations 2 and 3: Tables 8.2 and 8.3

Table 8.2: Constant instantaneous volatility and p(1)= -0.32 with noise to signal = 0.92: Annualised volatility: 29.5161 %			
	mean	MSE	Var
micr.RV	29.5372	0.0713	0.0716
DST	28.2334	1.8505	0.2072
RV	47.9422	369.4281	30.2076
RV.30min	47.9396	368.8906	29.7636

Table 8.3: Constant instantaneous volatility and p(1)= -0.48 with noise to signal = 3.5: Annualised volatility: 29.5161 %			
	mean	MSE	Var
micr.RV	29.9199	5.6545	5.5470
DST	29.1315	0.5399	0.3959
RV	155.2504	16408.39	605.3382
RV.30min	155.1981	16392.86	602.9326

The returns have first order autocorrelation with constant instantaneous volatility. We notice if the noise to signal ratio and the auto correlation are very high, the DST measurement performs best, but this is in an unrealistic environment. If the noise to signal ratio decreases, the microstructure realised volatility quickly becomes the best estimator Even at a very high noise to signal ratio and with an autocorrelation of 0.92 and -0.32 respectively, the microstructure realised volatility performs the best by far.

Simulation 4: Table 8.4

Table 8.4: Autocorrelation and changing instantaneous volatility Annualised volatility: 100.3243%			
	mean	MSE	Var
micr.RV	100.3488	0.5413	0.6693
DST	90.8785	97.9993	9.0542
RV	119.0969	996.8801	651.9170
RV.30min	119.1650	992.2786	644.6082

A changing instantaneous volatility is assumed. In this simulation a high annualised volatility, seldom seen in real life, is assumed to show more clearly how the other measurements break down under non-constant instantaneous volatility and under microstructure effects

Simulation 5: Table 8.5

Table 8.5: Constant instantaneous volatility and			
p(1)= -0.36, p(2)=0.124 with noise to signal = 0.92:			
Annualised volatility: 29.5161 %			
	mean	MSE	Var
micr.RV	29.5086	0.0326	0.0326
DST	28.2386	1.7895	0.1596
RV	46.3981	304.2548	19.5824
RV.30min	47.9396	304.5106	19.7060

Non-zero second order autocorrelation of returns is assumed. In the presence of second order autocorrelation the microstructure realised volatility again performs the best. The DST measurement is fairly robust against model misspecifications.

The microstructure realised volatility is the only measurement that performs well in all the simulations. It is also the only estimator which is unbiased and which has a small mean square error. In real life return data, where volatility changes and where microstructures are present, the microstructure realised volatility is the only measurement that still gives satisfactory results.

9

Final remarks and conclusions

Return series are non-stationary. Consequently, obtaining estimates for the second moment is not that easy a task. This non-stationary nature of returns resulted in researchers seeing volatility as an unobservable process. The main focus was to concentrate on ex-ante volatility models rather than on ex-post volatility measurements. This did not work successfully and researchers realised without a good volatility measurement, construction of good volatility models for forecasting purposes is almost impossible.

During the 1990's researchers shifted their attention to ex-post volatility by using non-parametric approaches. In the last few years many attempts have been made to define a volatility measurement that can handle all the characteristics present in real life return data. The two problematic characteristics of volatility of high frequency returns are non-zero autocorrelations of lag greater or equal to one and changes in the instantaneous volatility. While some volatility measurements capture one of the two characteristics quite well, no previous measurement has been able to handle both. The microstructure realised volatility proposed in this thesis is the first volatility measurement successful in doing this.

We have demonstrated the derivation of a microstructure realised volatility measurement that can handle first order autocorrelation and have mentioned that by adding the correct terms, the model can easily be extended to handle any autocorrelation lag. In the simulation, the dominance this measurement has over previous measurements under realistic situations, has been shown. The main reason for obtaining a good volatility measurement is for forecasting purposes, and subsequently a study to gauge how well this proposed measurement does under certain volatility models needs to be done.

References

[1] Anderson, T.G. and Bollershev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts, *International Economic Review* **39**, 885-905.

[2] Anderson, T.G., Borrerslev T. and Diebold, F.X (2003). Some Like it Smooth, and Some Like it Rough: Untangling Continous and Jump Components in Measuring, Modeling, and Forecasting Asset Return Volatility, Center for Financial Studies.

[3] Anderson, T.G., Borrerslev T., Diebold, F.X. and Labys, P. (2001a). The distribution of exchange rate volatility, *Journal of the American Statistical Association* **96**, 42-55

[4] Anderson, T.G., Borrerslev T., Diebold, F.X. and Labys, P. (2001b). Modelling and forecasting realised volatility, NBER Working Paper 8160. The National Bureau of Economic Research, Cambridge, MA.

[5] Barndoff-Nielsen, O.E. and Shepard, N (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics, *Journal of the Royal Statistical Society* **63**, 167-241.

[6] Barndoff-Nielsen, O.E. and Shepard, N. (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society* **64**, 253-280.

[7] Barndoff-Nielsen, O.E. and Shepard, N. (2002b). Estimating quadratic variation using realized volatility, *Journal of Applied Econometrics*.

[8] Barndoff-Nielsen, O.E. and Shepard, N. (2002c). How accurate is the asymptotic approximation to the distribution of realized volatility?, in D.W.F. Andrews, J.L. Powell, P.A. Ruud and J.H. Stock (eds), *Identification and Inference for Econometric models*.

[9] Barndoff-Nielsen, O.E. and Shepard, N. (2003). Power and Bipower variation with stochastic volatility and jumps, Manuscript Oxford University.

[10] Black, M. (1968). Studies in stock price volatility changes, Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section, *American Statistical Association*, 177-81.

[11] Cohen, K., S., M., R., S. and D., W. (1981). Transaction costs, order placement strategy, and existence of the bid-ask spread, *Journal of Political Economy* **89**, 287-305.

[12] Comte, F. and Renault, E. (2001). Long memory in continuous time stochastic volatility models, *Mathematical Finance* **8**,3291-323.

[13] Corsi, F. and Cursi, G. (2003). A Discrete Sine Transform Approach for Realized Volatility Measurement, Manuscript, University of Pisa.

[14] Corsi, F. (2003). A Simple Long Memory Model of Realised Volatility, Manuscript, University of Southern Switzerland.

[15] Dorogovcev, A.J. (1976). The consistency of an estimate of a parameter of a stochastic differential equation, *Theory Probab. Math. Statist*. **10**, 73-82.

[16] Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance in the UK inflation, *Econometrica* **50**, 987-1008.

[17] Fama, E. (1965). The behaviour of stock market prices, *Journal of Business* **38**, 34-105.

[18] Glosten, L. (1994). Is the electronic open limit-order book inevitable?, *Journal of Finance* **49**, 1127-1161.

[19] Glottlieb, G. and Kalay, A. (1985). Implication of the discreteness of observed stock prices, *Journal of Finance* **40**, 135-153.

[20] Hamilton, J.D., *Time series Analysis (1994)*. pp.281-283

[21] Hansen L.P. (1982). Large sample properties of the generalized method of moments, *Econometrica* **50**, 1029-1054.

[22] Hasbrouck, J. (1993). Assessing the quality of a security market: a new approach to transaction-cost measurement, *Review of Financial Studies* **6(1)**, 191-212

[23] Hasbrouck, J. (1996). Modelling market microstructure time series, in G.S. Maddala and C.R. Rao (eds), Statistical Methods in Finance, **Vol. 14** of *Handbook of Statistics*, North Holland, Amsterdam, pp. 647-692.

[24] Hull, J.C. (2003). *Options, Futures, and Other Derivatives, Fith Edition.*

[25] Jarrow, R. (1998). Volatility: New Estimation Techniques for Pricing Derivatives, *Risk Books*.

[26] Joyce, J. and Vogel, R. (1970). The uncertainty in risk: is variance unambiguous?, *Journal of Finance* **25**, 127-34.

[27] Kelly, L., Platen, E. and Sϕrensen, M. (2003). Estimation for Discretely Observed Diffusions using Transform Functions, Centre for Analytical Finance, *Working Paper Series* No.**151**.

[28] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations, Scand. J. *Statist*. **24**, 211-229.

[29] Neftci, S.N. (2000) Second Edition. *An Introduction to the Mathematics of Financial Derivatives*.

[30] Nelson, D. (1991). Conditional heteroscedasticity in asset pricing : A new approach. *Econometrica* **59**, 347-370)

[31] Merton, R.C. (1980). On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* **8**, 323-61

[32] Prakasa Rao, B.L.S. (1988). Statistical inference from sampled data for stochastic processes, *Contemp. Math.* **80**, 249-284.

[33] Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* **39**, 1127-1139.

Code for simulation

```
sim<-function(){
    normal.sample<-rnorm(72000)
    annualised.volatility<-0.0011*sqrt(var(normal.sample)*72000)
    log.price<-rep(0,72001)
    log.price[1]<-log(25)
    for(i in 1:72000){
        log.price[i+1]<-log.price[i]+0.0011*normal.sample[i]
    }
    bid.price<-(1/16)*floor(16*exp(log.price)-1)
    ask.price<-(1/16)*ceiling(16*exp(log.price)+1)
    bernoulli.vect<-rbinom(72001,1,0.5)
```

```
    observed.price<-bid.price*bernoulli.vect+ask.price*(1-bernoulli.vect)

    return.vect<-log(observed.price[-1]/observed.price[-length(observed.price)])

    mat.DST<-matrix(return.vect,ncol=40,byrow=T)

    mat.30.min<-matrix(return.vect,ncol=6,byrow=T)

    return.vect.30.min<-mat.30.min[,1]

    sigma.mat<-var(mat.DST)

    DST.volatility<-sqrt(eigen(sigma.mat)$values[40]*72000)

    realised.volatility.30.min<-sqrt(sum(return.vect.30.min^2))

    mod.realised.volatility<-sqrt(sum(return.vect^2)

    +2*sum(return.vect[-1]*return.vect[-length(return.vect)]))

    return(realised.volatility.30.min,mod.realised.volatility

    ,DST.volatility,annualised.volatility)

}

sim1<-function(B){

    DST<-rep(0,B)
```

```
    RV<-rep(0,B)
    for (i in 1:B){
        result<-sim()
        DST[i]<-result$DST.volatility
        RV[i]<-result$mod.realised.volatility
    }
    DST.mean<-mean(DST)
    RV.mean<-mean(RV)
    DST.err<-mean((DST-0.295160973)^2)
    RV.err<-mean((RV-0.295160973)^2)
    return(DST.mean,RV.mean,DST.err,RV.err)
}
sim1<-function(B){
    DST<-rep(0,B)
    RV<-rep(0,B)
    vol<-rep(0,B)
    for (i in 1:B){
        result<-sim.second()
```

```
        DST[i]<-result$DST.volatility
        RV[i]<-result$mod.realised.volatility
        vol[i]<-result$annualised.volatility
    }
    DST.mean<-mean(DST)
    RV.mean<-mean(RV)
    DST.err<-mean((DST-vol)^2)
    RV.err<-mean((RV-vol)^2)
    return(DST.mean,RV.mean,DST.err,RV.err)
}
sim<-function(sd){
    log.price<-rep(0,72001)
    log.price[1]<-log(100)
    instant.volatility<-rnorm(72000,110,sd)
    annualised.volatility<-sqrt(sum((instant.volatility*0.00001)^2))
    for(i in 1:72000){
        log.price[i+1]<-log.price[i]+instant.volatility[i]*rnorm(1,0,0.00001)
    }
```

```
bid.price<-(1/16)*floor(16*exp(log.price)-1)

ask.price<-(1/16)*ceiling(16*exp(log.price)+1)

bernoulli.vect<-rbinom(72001,1,0.5)

observed.price<-bid.price*bernoulli.vect+ask.price*(1-bernoulli.vect)

return.vect<-log(observed.price[-1]/observed.price[-length(observed.price)])

mat.DST<-matrix(return.vect,ncol=30,byrow=T)

mat.30.min<-matrix(return.vect,ncol=6,byrow=T)

return.vect.30.min<-mat.30.min[,1]

sigma.mat<-var(mat.DST)

DST.volatility<-sqrt(eigen(sigma.mat)$values[30]*72000)

realised.volatility.30.min<-sqrt(sum(return.vect.30.min^2))

mod.realised.volatility<-sqrt(sum(return.vect^2)

+2*sum(return.vect[-1]*return.vect[-length(return.vect)]))

return(annualised.volatility,realised.volatility.30.min

,mod.realised.volatility,DST.volatility)

}
```